Giáo trình Thống kê kinh doanh và kinh tế

Thống kê kinh doanh và kinh tế (Trường Đại học Kinh tế, Đại học Đà Nẵng)

Scan to open on Studeersnel

Studocu is not sponsored or endorsed by any college or university
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Chương 1

GIỚI THIỆU VỀ THỐNG KÊ

1.1. Thống kê và các ứng dụng

1.1.1. Thống kê và các phân nhánh

Thống kê là khoa học và nghệ thuật về thu thập, phân tích, trình bày và diễn giải dữ liệu về các hiện tượng số lớn nhằm trích xuất các thông tin hữu ích hỗ trợ việc ra các quyết định quản lý một cách có hiệu quả.

Thống kê mô tả là sử dụng các phương pháp như lập bảng, trình bày đồ thị hay xác định các con số thống kê đặc trưng để tóm lược dữ liệu theo cách trích rút được các thông tin hữu ích về hiện tượng nghiên cứu.

Thống kê suy diễn là quá trình sử dụng dữ liệu từ một bộ phận nhỏ của hiện tượng nghiên cứu (mẫu) để thực hiện các các phương pháp ước lượng và kiểm định các giả thuyết đặt ra cho toàn bộ hiện tượng nghiên cứu (tổng thể).

Thuật ngữ thống kê nói trên thường dùng để nói về ngành học hay lĩnh vực học thuật thống kê. Tuy nhiên, đôi khi thuật ngữ thống kê cũng được dùng để chỉ những con số thu thập được từ hoạt động thống kê mô tả hay thống kê suy diễn trên thực tế.

1.1.2. Các ứng dụng của thống kê

Các kỹ thuật thống kê được sử dụng rộng rãi bởi các nhà tiếp thị, các kế toán và kiểm toán viên, các nhà quản lý chất lượng, những người tiêu dùng, các nhà thể thao chuyên nghiệp, các nhà quản lý bệnh viện, các nhà giáo dục, các chính trị gia, các thầy thuốc, và nhiều người khác. Chẳng hạn:

Trong kế toán, các công ty kiểm toán sử dụng thủ tục lấy mẫu thống kê khi tiến hành kiểm toán cho khách hàng của họ.

Trong lĩnh vực tài chính, các nhà phân tích tài chính sử dụng một loạt các thông tin thống kê như tỷ lệ giá-lợi nhuận, suất cổ tức để đưa ra các khuyến nghị đầu tư của họ.

Trong tiếp thị, các máy quét điện tử tại các quầy thanh toán bán lẻ đang được sử dụng để thu thập dữ liệu cho một loạt các ứng dụng nghiên cứu thị trường.

Trong sản xuất, một loạt các biểu đồ kiểm soát chất lượng thống kê được sử dụng để giám sát các đầu ra của một quá trình sản xuất.

Trong kinh tế, các nhà kinh tế sử dụng thông tin thống kê trong việc đưa ra các dự báo về tương lai của nền kinh tế hoặc một số khía cạnh của nó.

1.2. Một số khái niệm cơ bản

1.2.1. Đơn vị tổng thể (phần tử), tổng thể, mẫu

Đơn vị tổng thể (phần tử) là một thực thể cấu thành hiện tượng nghiên cứu số lớn, trên đó các dữ liệu được thu thập làm cơ sở cho việc nghiên cứu thống kê về hiện tượng. Tuỳ mục đích nghiên cứu mà đơn vị tổng thể (phần tử) có thể khác nhau trên cùng một hiện tượng.

Ví dụ, khi nghiên cứu thống kê về các khiếm khuyết của một loại sản phẩm, một đơn vị tổng thể (phần tử) là một sản phẩm. Khi nghiên cứu về đơn thư khiếu nại của khách hàng về sản phẩm, một đơn thư khiếu nại là một đơn vị tổng thể (phần tử). Khi nghiên cứu thị hiếu của khách hàng về một loại sản phẩm, một khách hàng là một đơn vị tổng thể (phần tử).

Tổng thể là tập hợp tất cả các đơn vị tổng thể (phần tử) có chung các đặc điểm xác định một hiện tượng nghiên cứu cụ thể. Có hiện tượng, các đơn vị tổng thể (phần tử) biểu hiện rõ ràng, đầy đủ. Tổng thể này gọi là tổng thể bộc lộ. Ví dụ, tổng thể các sản phẩm trong một kho hàng. Có hiện tượng, ranh giới của tổng thể là không rõ ràng, không thể biết hết tất cả các đơn vị tổng thể (phần tử) mà chỉ biết các đặc tính qui định tổng thể đó. Tổng thể này được gọi là tổng thể tiềm ẩn. Ví dụ, tổng thể khách hàng có thể có của một loại sản phẩm.

Mẫu là một tập con của tổng thể. Mẫu thường bao gồm một nhóm nhỏ các đơn vị tổng thể (phần tử) được chọn đại diện cho tổng thể theo một phương pháp ngẫu nhiên nào đó.

1.2.2. Tiêu thức (biến), quan sát

Tiêu thức (biến) là khái niệm chỉ một đặc tính nào đó trên đơn vị tổng thể (phần tử) được chọn làm cơ sở để thu thập dữ liệu và nhận thức hiện tượng nghiên cứu.

Ví dụ, với tổng thể khách hàng có thể có của một loại sản phẩm, các tiêu thức (biến) có thể được chọn là: nghề nghiệp, sở thích, giới tính, độ tuổi, mức thu nhập, mức độ ưa thích sản phẩm…

Quan sát là khái niệm chỉ các giá trị (số đo) thu thập được trên các tiêu thức (biến) của một đơn vị tổng thể (phần tử). Tập hợp các các giá trị (số đo) thu thập được trên một đơn vị tổng thể (phần tử) được gọi là một quan sát.

Tiêu thức (biến) thường được phân biệt thành hai loại là định tính (thuộc tính) và định lượng (số lượng).

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Tiêu thức (biến) định tính là tiêu thức (biến) mà các quan sát của nó là các loại hình, các tính chất hoặc các con số định danh. Ví dụ, nghề nghiệp, sở thích, giới tính, số nhà…

Tiêu thức (biến) định lượng là tiêu thức (biến) mà các quan sát của nó là các con số định lượng (gọi là lượng biến). Ví dụ, độ tuổi, mức thu nhập… *Có hai loại lượng biến là lượng biến rời rạc và lượng biến liên tục.*

+ *Lượng biến rời rạc là* lượng biến chỉ nhận những giá trị nguyên. Ví dụ, số thành viên trong hộ, số xe máy sở hữu...

+ *Lượng biến liên tục là* lượng biến có khả năng nhận mọi giá trị trên trục số. Ví dụ, mức thu nhập, tiền lương, chi phí sản xuất...

1.2.3. Tham số tổng thể, thống kê mẫu

Một tham số tổng thể là một trị số tổng hợp của cả tổng thể nghiên cứu. *Một thống kê mẫu* là một trị số tổng hợp của một mẫu của tổng thể. Tham số tổng thể và thống kê mẫu được gọi chung là chỉ tiêu thống kê.

Ví dụ: Dữ liệu mẫu ở 1 thị trường năm 2014 cho biết: Tỉ lệ người có thu nhập trên 5 triệu đồng là 40% là một thống kê mẫu; dữ liệu từ một điều tra toàn bộ dân số của một nước cho biết: Thu nhập bình quân đầu người là 1,8 triệu là một tham số tổng thể.

1.3 Các loại thang đo trong thống kê

Tuỳ theo mức độ chặt chẽ của việc đo lường, người ta thường chia thang đo trong dữ liệu thống kê làm bốn loại sau:

- *Thang đo danh định:* Dữ liệu trên thang đo này chỉ thể hiện danh tính, không làm được bất kỳ phép tính nào từ so sánh đến cộng, trừ, nhân, chia. Ví dụ, giới tính, số nhà, số xe…Thang đo này thường dùng với tiêu thức định tính.

- *Thang đo thứ bậc:* Dữ liệu trên thang đo này thể hiện thứ bậc hơn, kém, cao, thấp nhưng với khoảng cách (đơn vị) không đều. Ví dụ, mức độ ưa thích một loại sản phẩm… Dữ liệu trên thang đo này chỉ làm được phép tính so sánh. Các phép toán khác không bảo đảm ý nghĩa. Thang đo này thường được dùng với tiêu thức định tính.

- *Thang đo khoảng:* Dữ liệu trên thang đo này thể hiện rõ độ lớn hơn kém với khoảng cách (đơn vị) đều nhưng không có số không tuyệt đối. Ví dụ, điểm ưa thích sản phẩm được cho trên thang đo điểm 10, nhiệt độ ... Thang đo này được dùng với tiêu thức định lượng. Quan hệ tỉ lệ giữa các con số trên thang đo này không bảo đảm ý nghĩa vì không có số không tuyệt đối.

- *Thang đo tỉ lệ:* Dữ liệu trên thang đo này thể hiện rõ độ lớn hơn, kém với khoảng cách (đơn vị) đều và có số không tuyệt đối. Ví dụ, mức thu nhập, số khuyết tật của sản phẩm…Thang đo này được dùng với tiêu thức định lượng. Dữ liệu trên thang đo này làm được mọi phép tính với đầy đủ ý nghĩa.

Để nâng cao độ chính xác của đo lường và khả năng vận dụng các phương pháp phân tích thống kê, khi thu thập dữ liệu cần chú ý sử dụng tối đa, có thể được, các thang đo định lượng.

1.4. Dữ liệu dùng trong thống kê

Dữ liệu là những sự kiện và con số được thu thập, phân tích và tóm lược nhằm trình bày và giải thích về các hiện tượng nghiên cứu. Tất cả dữ liệu được thu thập cho một nghiên cứu cụ thể được gọi là tập hợp dữ liệu cho nghiên cứu đó.

1.4.1. Dữ liệu tổng thể, dữ liệu mẫu

- *Dữ liệu tổng thể:* Là dữ liệu được thu thập trên tất cả các đơn vị tổng thể.

Loại dữ liệu này cho phép tổng hợp trực tiếp các chỉ tiêu về toàn bộ tổng thể. Tuy nhiên chi phí của nó thường rất cao nên thống kê kinh doanh ít dùng loại dữ liệu này.

- *Dữ liệu mẫu:* Là dữ liệu được thu thập trên tập con các đơn vị tổng thể được chọn đại diện cho tổng thể.

Dữ liệu mẫu giúp giảm thiểu rất nhiều chi phí và thời gian nghiên cứu nên rất hay được dùng trong thống kê kinh doanh và kinh tế.

1.4.2. Dữ liệu chéo, dữ liệu chuỗi thời gian

- *Dữ liệu chéo:* Là dữ liệu thu thập theo từng đơn vị tổng thể, tại một thời gian nhất định. Trên từng đơn vị tổng thể, dữ liệu được thu thập theo một số tiêu thức được chọn lựa phục vụ cho việc nghiên cứu hiện tượng.

Dạng tổng quát của dữ liệu chéo như sau:

Đơn vị tổng thể	Tiêu thức 1	Tiêu thức 2	…	Tiêu thức k
1	x_{11}	x_{12}		x_{1k}
2	x_{21}	x_{22}		x_{2k}
3	x_{31}	x_{32}		x_{3k}
…	…	…		…
n	x_{n1}	x_{n2}		x_{nk}

Trong bảng trên, dữ liệu của một đơn vị tổng thể (phần tử) được thể hiện trên một dòng đó chính là một quan sát.

Ví dụ: Có dữ liệu mẫu về 18 đơn thư khiếu nại của khách hàng được chọn ngẫu nhiên.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Số thứ tự đơn thư	Tuổi của khách hàng	Giới tính của khách hàng	Nơi mua	Gía trị sản phẩm (tr.đ)	Lần khiếu nại thứ	Thời gian gặp sự cố (ngày)	Loại sự cố	Yêu cầu của khách hàng
1	22	Nữ	A	2,5	1	32	Kêu	B.T
2	26	Nam	B	1,8	2	24	Bể	Đổi
3	25	Nam	A	12,5	1	67	Rỉ	Đổi
4	27	Nữ	B	4,5	3	33	Nứt	B.T
5	26	Nữ	B	2,8	1	28	Cháy	B.T
6	26	Nữ	D	6,4	2	64	Cháy	B.T
7	25	Nam	F	10,2	1	45	Kêu	Sửa
8	27	Nữ	A	3,5	2	21	Nứt	Đổi
9	26	Nam	C	6,8	1	29	Nứt	B.T
10	48	Nữ	C	5,5	3	21	Rỉ	Đổi
11	26	Nam	A	4,7	2	12	Bể	Sửa
12	25	Nam	A	8,2	1	48	Kêu	Đổi
13	26	Nam	C	9,1	2	57	Rỉ	Sửa
14	25	Nữ	B	7,4	2	42	Bể	B.T
15	60	Nam	A	5,8	1	22	Cháy	Sửa
16	27	Nữ	B	4,4	2	34	Nứt	B.T
17	26	Nữ	B	9,7	1	68	Bể	B.T
18	27	Nam	C	2,6	1	39	Kêu	Sửa

Kí hiệu: B.T: Bồi thường

- *Dữ liệu chuỗi thời gian:* Là dữ liệu về một hiện tượng nghiên cứu được thu thập ở nhiều thời gian khác nhau.

Ví dụ: Có dữ liệu về lợi nhuận của một doanh nghiệp như sau.

Năm	2004	2005	2006	2007	2008	2009	2010	2011
Lợi nhuận (tr.đ)	300	250	400	500	800	700	900	1200

1.5. Các nguồn dữ liệu sử dụng trong thống kê

Có hai nguồn dữ liệu được sử dụng trong thống kê là dữ liệu sơ cấp và dữ liệu thứ cấp.

- *Dữ liệu sơ cấp:* Là loại dữ liệu do đơn vị nghiên cứu tổ chức thu thập trực tiếp từ đối tượng nghiên cứu hoặc thuê một tổ chức chuyên nghiệp khác thu thập.

Việc thu thập dữ liệu sơ cấp có thể được tiến hành thông qua điều tra (quan sát) thống kê trên toàn bộ các đơn vị tổng thể của tổng thể gọi là điều tra toàn bộ hoặc điều tra (quan sát) thống kê trên một mẫu của tổng thể gọi là điều tra mẫu.

Việc thu thập dữ liệu sơ cấp cũng có thể được tiến hành thông qua nghiên cứu thực nghiệm. Trong một nghiên cứu thực nghiệm, một số biến quan tâm được xác định trước. Sau đó một hoặc nhiều biến khác được xác định, điều chỉnh hoặc kiểm soát sao cho dữ liệu thu được phản ánh được ảnh hưởng của chúng đến biến quan tâm ban đầu như thế nào.

Nguồn dữ liệu sơ cấp có độ chính xác cao, bảo đảm tính cập nhật nhưng tốn nhiều thời gian và chi phí.

- *Dữ liệu thứ cấp:* Là loại dữ liệu được thu thập từ các nguồn tài liệu có sẵn bên trong hay bên ngoài doanh nghiệp như các chứng từ sổ sách của doanh nghiệp, các tập san, tạp chí chuyên đề, niên giám thống kê của tổng cục thống kê, các công trình nghiên cứu đã công bố, dữ liệu của IMF, dữ liệu của Wordbank, dữ liệu trên mạng internet, ...

Ví dụ: Dữ liệu sẵn có từ các hồ sơ nội bộ các công ty

Nguồn	Một số dữ liệu sẵn có
Hồ sơ nhân viên	Tên, địa chỉ, số an sinh xã hội
Hồ sơ sản xuất	Số bộ phận, số lượng sản xuất, chi phí nhân công trực tiếp, chi phí nguyên liệu
Hồ sơ tồn kho	Số bộ phận, số lượng tồn kho, mức đặt hàng lại, số lượng đơn hàng
Hồ sơ bán hàng	Số sản phẩm, lượng hàng bán, lượng hàng bán theo vùng
Hồ sơ tín dụng	Tên khách hàng, hạn mức tín dụng, khoản phải thu
Hồ sơ khách hàng	Tuổi, giới tính, thu nhập, số người trong hộ

Nguồn dữ liệu thứ cấp ít tốn thời gian và chi phí thu thập nhưng thường thiếu tính cập nhật, kém phù hợp, đôi khi không đầy đủ.

Khi sử dụng dữ liệu trong thống kê cần lưu ý sai số. Sai số dữ liệu là chênh lệch giữa các giá trị thu thập được và giá trị thực tế của hiện tượng. Sử dụng dữ liệu có sai số lớn còn nguy hại hơn cả không có dữ liệu để dùng.

1.6. Đạo dức nghề nghiệp trong thực hành thống kê

Các vấn đề đạo đức nảy sinh trong thống kê bởi vì vai trò của thống kê trong thu thập, phân tích, trình bày và diễn giải dữ liệu.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Trong thống kê, hành vi vô đạo đức có thể bao gồm nhiều dạng như lấy mẫu thiên lệch, phân tích dữ liệu không thích hợp, vẽ biểu đồ gây hiểu nhầm, sử dụng các thống kê mô tả không thích hợp hay diễn giải thiên lệch các kết quả thống kê.

Vì vậy, khi thực hành thống kê phải công bằng, kỹ lưỡng, khách quan và trung lập trong thu thập dữ liệu, tiến hành phân tích, trình bày và viết báo cáo nghiên cứu.

1.7. Một số phần mềm phân tích thống kê

Công việc phân tích dữ liệu thống kê số lớn rất phức tạp và nặng nề có thể được hỗ trợ một cách rất đắc lực của các phần mềm phân tích thống kê. Đặc biệt tiện dụng trong số này là phần mềm phân tích thống kê chuyên nghiệp **SPSS**. Có thể nói sau khi nhập liệu vào máy tính, phần mềm này cho phép sử dụng hầu hết các phương pháp phân tích thống kê để khai thác dữ liệu đã được nhập một cách tự động, nhanh chóng với kết quả kết xuất rất rõ ràng và đầy thuyết phục. Việc kết xuất các kết quả phân tích từ phần mềm **SPSS** để lập các báo cáo phân tích trên **Word** hay **PowerPoint** cũng rất tiện lợi và nhanh chóng.

Nếu chỉ khai thác riêng lẻ dữ liệu trên một vài phương pháp phân tích thống kê giới hạn nào đó có thể sử dụng phần mềm thông dụng **Excel** cũng rất tiện lợi và nhanh chóng.

Chương 2 THỐNG KÊ MÔ TẢ

Dữ liệu chéo mới thu thập được thường rất nhiều và rối rắm. Chúng ta thường bị nhiễu loạn và rất khó nhận thức được điều gì hữu ích về hiện tượng nghiên cứu trước một khối lượng lớn dữ liệu như vậy. *Các phương pháp thống kê mô tả dữ liệu chéo giúp tóm lược dữ liệu nhằm làm bộc lộ các đặc trưng cơ bản nhất, đáng quan tâm nhất về hiện tượng nghiên cứu. Mục đích là cung cấp cái nhìn sâu hơn về dữ liệu mà chúng ta không thể thấy được ngay trên dữ liệu ban đầu.*

2.1. Mô tả (tóm tắt) dữ liệu cho một tiêu thức (biến) định tính bằng bảng phân phối và biểu đồ

2.1.1. Lập bảng phân phối

- Trường hợp tiêu thức (biến) có ít biểu hiện kiểu loại khác nhau: Bảng phân phối tần số được lập với một kiểu loại thuộc tính khác nhau là một tổ (nhóm).

Ví dụ: Xét dữ liệu chéo ở trang 3 về 18 đơn thư khiếu nại. Bảng phân phối tần số theo tiêu thức (biến) yêu cầu của khách hàng được lập như sau. Trong đó, tần số là số đếm các quan sát trong mỗi tổ (nhóm) yêu cầu của khách hàng.

Yêu cầu của khách hàng	Tần số
Sửa	5
Đổi	5
Bồi thường	8

Bảng phân phối tần số trên cho thấy các yêu cầu của khách hàng khiếu nại xuất hiện khá đều trên cả ba loại yêu cầu, trong đó yêu cầu bồi thường có phần nhiều hơn.

Ngoài phân phối tần số như trên, bảng phân phối có thể được lập theo phân phối tần suất hay tần suất phần trăm. *Tần số* (f_i) của một tổ (nhóm) là số quan sát trong tổ (nhóm) đó. *Tần suất* là tỉ trọng hay tỉ lệ (f_i/n) giữa tần số (f_i) so với tổng số quan sát (n) của dữ liệu. *Tần suất phần trăm* bằng tần suất nhân với 100. Nó cho biết mỗi tổ (nhóm) chiếm bao nhiêu phần trăm trên mẫu hay tổng thể.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

- Trường hợp tiêu thức (biến) có nhiều biểu hiện kiểu loại khác nhau:

Để tránh hiện tượng bảng phân phối được lập với quá nhiều kiểu loại có tần số rất thấp làm cho bảng quá dài và khó nhận thức, người ta thường ghép các biểu hiện kiểu loại gần giống nhau về tính chất thành một một số tổ (nhóm không chồng lẫn) sao cho thuận lợi trong việc nhận thức bản chất của hiện tượng. Có thể thử một vài cách ghép khác nhau để từ đó chọn ra cách ghép cho nhận thức rõ nhất về hiện tượng.

Ví dụ, Xét dữ liệu chéo ở trang 3. Bảng phân phối tần số theo tiêu thức (biến) loại sự cố có thể được lập như sau.

Tiêu thức (biến) loại sự cố ở đây thực sự có không quá nhiều loại sự cố khác nhau (6 loại). Bảng phân phối có thể được lập với 6 tổ (nhóm) là 6 loại sự cố khác nhau: bể, nứt, rỉ, cháy, hỏng, kêu.

Tuy nhiên, giả sử 6 loại sự cố được xem là khá nhiều so với 18 đơn thư khiếu nại và không hữu ích lắm trong quản lý sự cố. Bảng phân phối cũng có thể được lập bằng cách ghép 6 loại sự cố thành 2 tổ (nhóm): tổ (nhóm) thứ nhất với tên gọi là sự cố vật liệu bao gồm 3 loại sự cố bể, nứt, rỉ; tổ (nhóm) thứ hai với tên gọi là sự cố kỹ thuật bao gồm 3 loại sự cố cháy, hỏng, kêu.

Loại sự cố	Tần số
Sự cố vật liệu	10
Sự cố kỹ thuật	8

Bảng phân phối tần số trên cho thấy hai loại sự cố vật liệu và kỹ thuật xuất hiện khá đều nhau, trong đó sự cố vật liệu có phần nhiều hơn.

2.1.2. Trình bày bằng biểu đồ

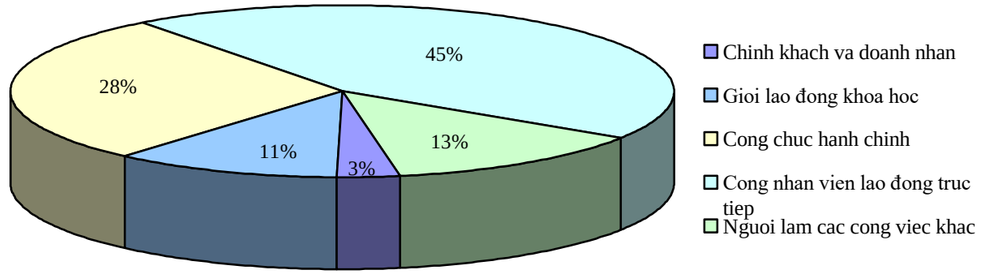
Phân phối của tiêu thức (biến) định tính thường được mô tả (trình bày) bằng biểu đồ hình bánh (hình tròn) hay biểu đồ hình thanh. *Trình bày dữ liệu trong bảng phân phối lên đồ thị thích hợp sẽ giúp ta mô tả tóm tắt các đặc trưng phân phối của hiện tượng nghiên cứu bằng hình ảnh.*

Ví dụ, Có bảng phân phối 900 người tiêu dùng theo các nhóm nghề nghiệp:

Nhóm nghề nghiệp	Tần số (f_i)	Tần suất %
- Chính khách và doanh nhân	30	3
- Giới lao động khoa học	100	11
- Công chức hành chính	250	28
- Công nhân viên lao động trực tiếp	400	45
- Người làm các công việc khác	120	13
Tổng cộng	900	100

Người ta thường mô tả phân phối tần suất bằng biểu đồ hình bánh. Trong đó, 1% của tần suất phần trăm tương đương với 3,6 độ ở tâm hình tròn.

BIỂU ĐỒ HÌNH BÁNH (PIE CHARTS)

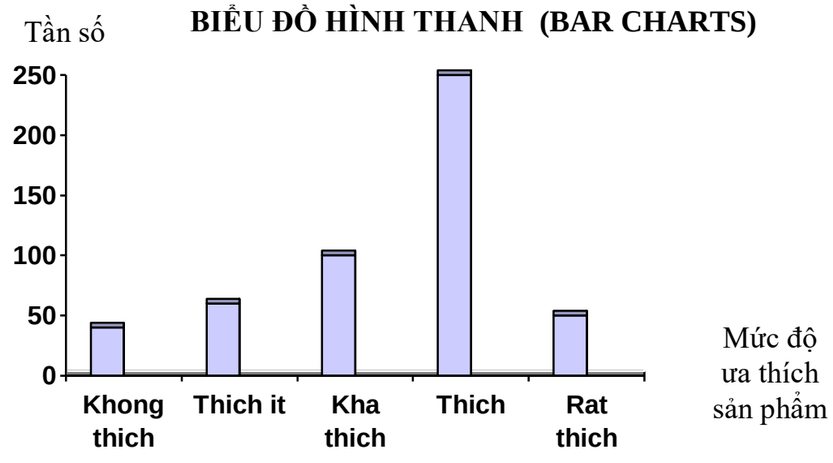


Dữ liệu trong bảng phân phối tần số của tiêu thức (biến) định tính thường được trình bày trên biểu đồ hình thanh.

Ví dụ, Có bảng phân phối 500 người tiêu dùng theo mức độ ưa thích sản phẩm như sau:

Mức độ ưa thích sản phẩm	Tần số (f_i)	Tần suất (%)
- Không thích	40	8
- Thích ít	60	12
- Khá thích	100	20
- Thích	250	50
- Rất thích	50	10
Tổng cộng	500	100

BIỂU ĐỒ HÌNH THANH (BAR CHARTS)



Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

2.2. Mô tả (tóm tắt) dữ liệu cho một tiêu thức (biến) định lượng bằng bảng phân phối và biểu đồ

2.2.1. Lập bảng phân phối

- Trường hợp tiêu thức (biến) là rời rạc và biến thiên ít: Bảng phân phối tần số được lập với mỗi giá trị rời rạc là một tổ (nhóm).

Ví dụ: Xét dữ liệu về 18 đơn thư khiếu nại ở trang 3. Bảng phân phối theo tiêu thức (biến) số lần khiếu nại của khách hàng được lập như sau.

Số lần khiếu nại	Tần số
1	9
2	7
3	2

Bảng phân phối tần số trên cho thấy số lần khiếu nại của khách hàng khiếu nại xuất hiện giảm dần theo số lần khiếu nại, trong đó tập trung nhiều nhất là khiếu nại lần đầu, tiếp đến là khiếu nại lần thứ hai.

- Trường hợp tiêu thức (biến) là liên tục hay rời rạc và biến thiên nhiều:

Trước hết cần phân tổ (phân nhóm) dữ liệu thành một số tổ (nhóm). Thông thường người ta chọn phân tổ đều với một số lượng tổ chọn trước. Mỗi tổ sẽ có hai giới hạn:

+ Giới hạn dưới: là lượng biến nhỏ nhất của tổ làm cho tổ đó hình thành.

+ Giới hạn trên: là lượng biến lớn nhất của tổ. Vượt quá giới hạn này sẽ sang tổ khác.

Chênh lệch giữa hai giới hạn mỗi tổ gọi là Trị số khoảng cách tổ.

Công thức tính trị số khoảng cách tổ đều:

$$h = (x_{max} - x_{min}) / k$$

Trong đó: h : Trị số khoảng cách tổ
x_{max} : Lượng biến lớn nhất
x_{min} : Lượng biến nhỏ nhất
k : Số tổ lựa chọn

Để đạt hiệu quả cao trong mô tả tóm tắt dữ liệu, người ta thường cân nhắc chọn k trong khoảng từ 5 đến 20. Nguyên tắc chung là số đơn vị tổng thể nhiều thì chọn k lớn và ngược lại. Có thể thử một vài giá trị của k để tìm giá trị k sao cho bức tranh phân phối rõ nhất, hữu ích nhất về hiện tượng.

Một công thức thống kê kinh nghiệm có thể tham khảo để xác định k:

$$k = (2 \times n)^{0,333}$$

Trong đó: k : Số tổ lựa chọn
n : Số đơn vị tổng thể

Ví dụ: Có dữ liệu mẫu về thu nhập (triệu/người) của 60 người tiêu dùng tại thị trường X như sau:

0,52	1,05	1,50	1,60	2,80	4,00
0,64	1,05	1,50	1,60	2,80	4,20
0,70	1,05	1,50	1,80	2,90	5,00
0,70	1,20	1,50	1,80	3,00	5,00
0,80	1,20	1,50	1,80	3,00	5,60
0,80	1,20	1,50	2,00	3,00	6,20
0,80	1,30	1,60	2,00	3,10	6,20
0,90	1,30	1,60	2,00	3,20	6,50
0,90	1,30	1,60	2,00	3,20	6,80
0,90	1,30	1,60	2,50	3,50	7,00

Chọn số tổ theo công thức $k = (2 \times n)^{0,333} = (2x60)^{0,333} = 5$

Trị số khoảng cách tổ : $h = (x_{max} - x_{min}) / k = (7,00 - 0,52) / 5 = 1,30$

h được xác định cùng một độ chính xác với dữ liệu (cùng số chữ số sau dấu phẩy) nhưng theo nguyên tắc làm tròn lên trên. Trường hợp chia chẵn đến độ chính xác này thì tăng thêm một đơn vị cho chữ số cuối cùng.

Bảng phân phối tần số:

Mức thu nhập (tr.đ)	Tần số
0,52 đến dưới 1,82	35
1,82 đến dưới 3,12	12
3,12 đến dưới 4,42	5
4,42 đến dưới 5,72	3
5,72 đến dưới 7,02	5

Tần số của mỗi tổ được xác định bằng cách đếm số quan sát nằm trong khoảng giá trị của mỗi tổ. Trường hợp quan sát trùng với giới hạn tổ, đếm lượng biến đó vào tổ lớn hơn.

Bảng phân phối tần số trên cho thấy người tiêu dùng tập trung chủ yếu ở hai nhóm thu nhập thấp nhất, các nhóm thu nhập trung bình và cao chiếm phần rất ít.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Ngoài phân phối tần số như trên, bảng phân phối có thể được lập theo phân phối tần suất hay tần suất phần trăm. *Tần suất* là tỉ trọng hay tỉ lệ (f_i/n) giữa tần số (f_i) so với tổng số quan sát (n) của dữ liệu. *Tần suất phần trăm* bằng tần suất nhân với 100. Nó cho biết mỗi tổ (nhóm) chiếm bao nhiêu phần trăm trên mẫu hay tổng thể.

Chú ý: Một số kỹ thuật sau có thể được sử dụng trong phân tổ (phân nhóm).

a). *Phân tổ (phân nhóm) với các giới hạn tổ không trùng với các quan sát:*

Giới hạn dưới tổ đầu tiên được xác định nhỏ hơn x_{min} một nửa đơn vị của chữ số cuối cùng.

Ví dụ 1: Phân tổ thu nhập của 40 khách hàng với x_{min} = 4,23 triệu đồng và x_{max} = 10,32 triệu đồng thành 4 tổ.

$$(x_{max} - x_{min}) / k = (10,32 – 4,23) / 4 = 1,523$$

Chọn h = 1,53 và chọn giới hạn dưới tổ đầu tiên là 4,225

Các giới hạn tổ sẽ là:

Thu nhập (triệu đồng)
4,225 – 5,755
5,755 – 7,285
7,285 – 8,815
8,815 – 10,345

Ví dụ 2: Phân tổ thu nhập của 40 khách hàng với x_{min} = 3,2 triệu đồng và x_{max} = 9,6 triệu đồng thành 4 tổ.

$$(x_{max} - x_{min}) / k = (9,6 – 3,2) / 4 = 1,6$$

Chọn h =1,7 và chọn giới hạn dưới tổ đầu tiên là 3,15

Các giới hạn tổ sẽ là:

Thu nhập (triệu đồng)
3,15 – 4,85
4,85 – 6,55
6,55 – 8,25
8,25 – 9,95

b) Chọn các giới hạn tổ nguyên và không trùng nhau khi phân tổ với các lượng biến (quan sát) rời rạc:

Tiến hành tương tự trường hợp a. Sau đó, giới hạn dưới được làm tròn lên, giới hạn trên được làm tròn xuống. Cuối cùng, có thể dịch chuyển các giới hạn tổ về phía trái sao cho chúng cân xứng hơn với dữ liệu gốc.

Ví dụ 3: Phân tổ 40 công nhân trong một doanh nghiệp theo tuổi nghề với x_{min} = 12 và x_{max} = 33 thành 5 tổ.

$$(x_{max} - x_{min}) / k = (33 – 12) / 5 = 4,2$$

Như trường hợp a, chọn h = 5 và chọn giới hạn dưới tổ đầu tiên là 11,5.

Các giới hạn tổ được xác định qua các bước sau:

Tuổi nghề		Tuổi nghề		Tuổi nghề
11,5 – 16,5	Làm	12 – 16	Dịch	10 – 14
16,5 – 21,5	tròn	17 – 21	sang	15 – 19
21,5 – 26,5	=>	22 – 26	trái	20 – 24
26,5 – 31,5		27 – 31	=>	25 – 29
31,5 – 36,5		32 – 36		30 – 34

Ở ví dụ trên, giới hạn trên cùng sau khi làm tròn bị tràn qua phải so với dữ liệu gốc: 36-33=3 đơn vị. Do đó, có thể dịch các giới hạn tổ sau khi làm tròn qua trái 3/2 ≈ 2 đơn vị.

Trong trường hợp các giới hạn tổ không trùng nhau, trị số khoảng cách tổ có thể được tính bằng hiệu của hai giới hạn dưới của hai tổ kế nhau: h=30-25=25-20=20-15=15-10=5.

c) Phân tổ dựa trên sự khác nhau rõ rệt về tính chất giữa các tổ:

Trong một số trường hợp, người ta có thể dựa vào các môn khoa học khác, dựa vào thực nghiệm, hoặc dựa vào kinh nghiệm để xác định các giới hạn tổ sao cho các tổ có sự khác nhau rõ rệt về tính chất.

Ví dụ, một doanh nghiệp dựa vào thực nghiệm "nếm độ ngọt" để phân chia người tiêu dùng ở một thị trường theo độ tuổi có đặc tính ưa thích độ ngọt khác nhau thành các tổ sau:

Dưới 16 tuổi
16 - 25
26 - 45
46 - 60
Trên 60 tuổi

2.2.2. Mô tả (trình bày) bằng biểu đồ

Phân phối của tiêu thức (biến) định lượng thường được mô tả (trình bày) bằng các loại biểu đồ như biểu đồ điểm, biểu đồ phân phối, biểu đồ hình cung hay biểu đồ cành và lá. *Trình bày dữ liệu trong bảng phân phối lên biểu đồ thích hợp sẽ giúp ta mô tả tóm tắt các đặc trưng phân phối của hiện tượng nghiên cứu bằng hình ảnh.*

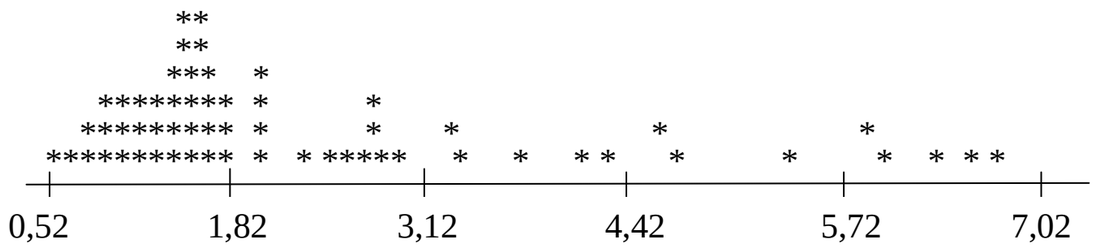
Biểu đồ điểm hoặc biểu đồ phân phối thường được dùng cho dữ liệu định lượng không có khoảng cách tổ hoặc khoảng cách tổ đều.

Dit document is beschikbaar op studeersnel

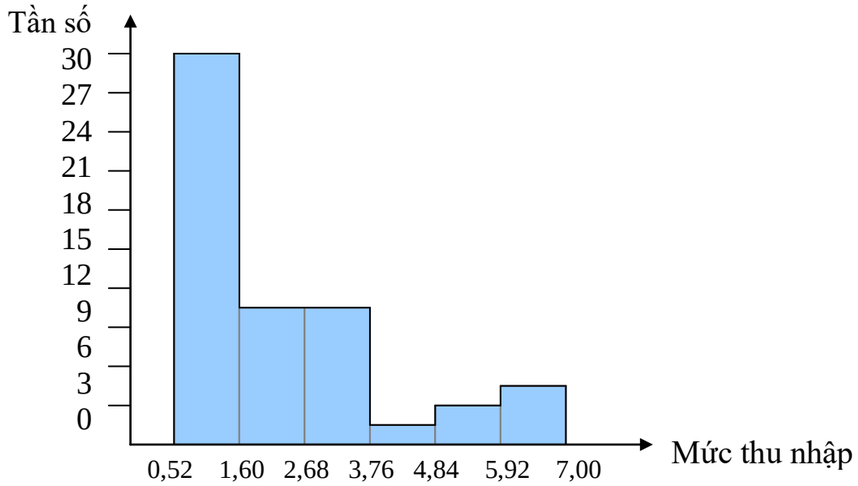
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Ví dụ: Dữ liệu mẫu về thu nhập của 60 người tiêu dùng tại thị trường X đã được lập bảng phân phối ở trên có thể được trình bày trên biểu đồ điểm hoặc biểu đồ phân phối như sau:

Biểu đồ điểm (Dot plot):



Biểu đồ phân phối (histogram):



Phân phối tích luỹ và biểu đồ hình cung: Ngoài phân phối tần số như trên, đôi khi người ta sử dụng phân phối tích luỹ.

Ví dụ, Có bảng phân phối về mức thu nhập của 60 người tiêu dùng.

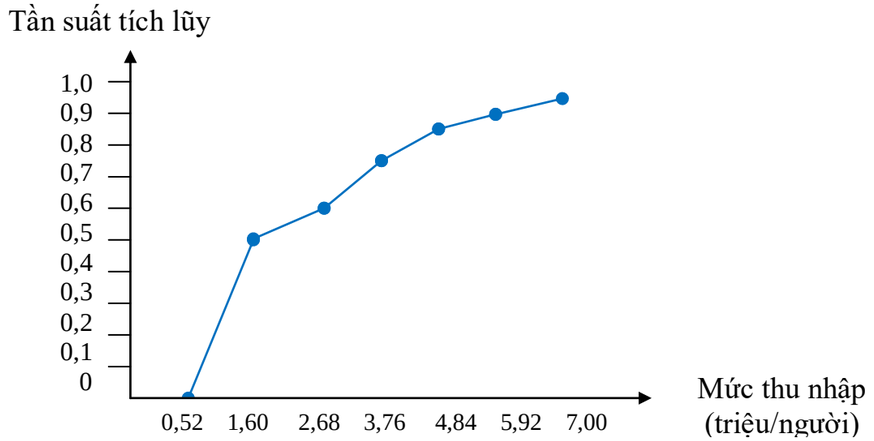
Mức thu nhập (triệu/người)	Tần số (f_i)	Tần suất (%)	Tần số tích luỹ (S_i)	Tần suất tích luỹ (%)
0,52 - 1,60	30	50	30	50
1,60 - 2,68	10	17	40	67
2,68 - 3,76	10	17	50	84
3,76 - 4,84	2	3	52	87
4,84 - 5,92	3	5	55	92
5,92 - 7,00	5	8	60	100

Tần số tích luỹ là số cộng dồn các tần số của các tổ kể từ tổ đầu tiên cho đến tổ đang xét.

Tần suất tích luỹ **%** là số cộng dồn các tần suất % của các tổ kể từ tổ đầu tiên cho đến tổ đang xét. Nó cho biết bộ phận gồm các tổ kể từ tổ đang xét cho đến tổ đầu tiên chiếm bao nhiêu phần trăm.

Bảng phân phối tích lũy trên cho thấy có đến 84% số người có mức lương dưới 3,76 triệu đồng.

Biểu đồ hình cung (ogive)



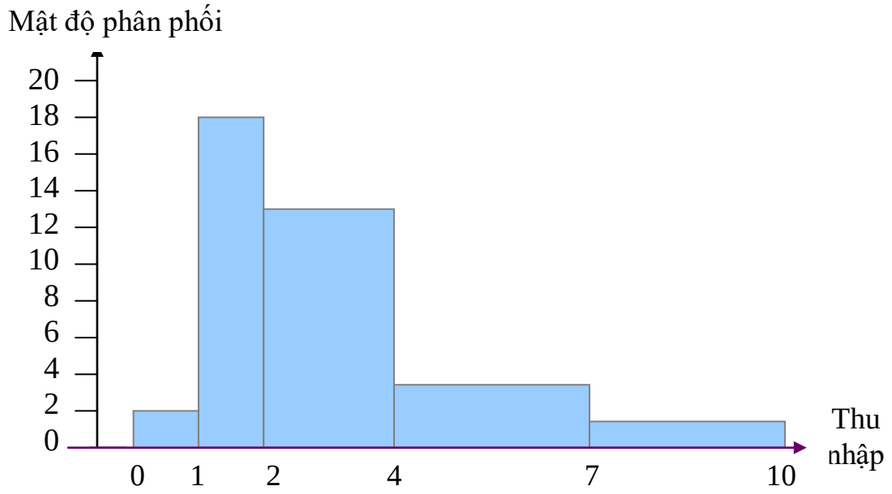
Biểu đồ mật độ phân phối: Dữ liệu trên bảng phân phối có khoảng cách tổ không đều thường không được mô tả bằng biểu đồ phân phối tần số. Bởi vì, nó không cho cảm nhận thị giác đúng về phân phối này. Do đó, trường hợp này, người ta thường vẽ biểu đồ phân phối theo mật độ phân phối là số quan sát tính trên một đơn vị khoảng cách tổ: $p_i = f_i / h_i$ với h_i là trị số khoảng cách tổ..

Ví dụ, Bảng phân phối tần số về thu nhập của 600 người tiêu dùng.

Thu nhập (triệu đồng)	Tần số (f_i)	Tần suất (%)	Mật độ phân phối (p_i)
Dưới 1	2	3	2,00
1 - 2	18	30	18,00
2 - 4	26	43	13,00
4 - 7	10	17	3,33
Trên 7	4	67	1,33

Biểu đồ phân phối của bảng phân phối trên phải được vẽ dựa trên mật độ phân phối như sau.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)



Biểu đồ cành và lá (Stem and leaf diagram)

Biểu đồ cành và lá là một cách trình bày tiêu thức (biến) định lượng một cách hình ảnh. Nó cho chúng ta sự nhận thức không những về sự biến thiên mà cả sự phân phối của biến định lượng.

Mỗi trị số của tiêu thức được chia làm hai phần cành và lá. Lá gồm một chữ số cuối cùng bên phải. Cành gồm các chữ số còn lại bên trái chữ số của lá (nếu không có thì lấy bằng 0).

Các trị số có cành giống nhau được sắp cùng một hàng (chung cành) nhưng phân biệt nhau bằng lá được sắp xếp theo thứ tự từ nhỏ đến lớn. Các cành được sắp xếp theo thứ tự từ nhỏ đến lớn (hoặc ngược lại). Mỗi cành cách nhau một đơn vị (hoặc hàng chục, hàng trăm…). Cành nào không có dữ liệu vẫn được ghi nhưng ở phần lá của nó thì để trống. Giữa cành và lá tách nhau bằng một đường thẳng đứng.

Ví dụ: Có dữ liệu về độ tuổi của 30 khách hàng như sau.

5 5 8 8 9 12 13 14 14 17 19 19 22 24 24 27
31 35 38 41 48 49 62 65 68 82 85 89 95 106

Trị số thứ nhất 5 có lá là 5, cành là 0. Trị số 12 có lá là 2, cành là 1. Trị số 106 có lá 6, cành 10. Không có trị số nào từ 50 đến 59 do đó cành 5 không có lá nào. Các trị số 5, 8, 8, 9 có chung cành là 0…

Để biểu đồ cành và lá đạt hiệu quả cao trong mô tả tóm tắt dữ liệu, số cành thường được giới hạn trong khoảng từ 5 đến 20.

Biểu đồ cành và lá của dữ liệu trên là như sau:

0	5 8 8 9
1	2 3 4 4 7 9 9
2	2 4 4 7
3	1 5 8
4	1 8 9
5	
6	2 5 8
7	
8	2 5 9
9	5
10	6

Khi trình bày biểu đồ cành và lá, ta không cần quan tâm đến dấu phẩy thập phân mà chỉ cần nói rõ đơn vị tính của lá.

Nếu dữ liệu biến thiên quá nhiều, có thể xây dựng biểu đồ cành và lá theo thủ thuật sau: trên dữ liệu không xét dấu thập phân, lấy trị số lớn nhất trừ trị số nhỏ nhất được một hiệu số. Bỏ bớt k chữ số bên phải của hiệu số này và làm tròn, sao cho giá trị còn lại nằm trong khoảng từ 20 đến 200. Lấy chữ số cuối làm lá, số cành sẽ nằm trong khoảng chừng từ 2 đến 20 cành.

Ví dụ, có dữ liệu của một tiêu thức thu nhập (triệu đồng) như sau:

2,8 11,2 34,8 62,5 102,0 105,6 452,8 503,2 668,3

Dữ liệu đã bỏ dấu thập phân:

28 112 348 625 1020 1056 4528 5032 6683

Chênh lệch giữa số lớn nhất và số nhỏ nhất: 6683 - 28 = 6655. Nếu bỏ đi hai chữ số bên phải, hiệu số này còn 66. Sử dụng chữ số cuối cùng làm lá, sẽ có 7 cành từ cành 0 đến cành 6. Số cành này nằm trong khoảng từ 5 đến 20. Vì vậy, bỏ 2 chữ số bên phải (dữ liệu đã bỏ dấu thập phân), ta được:

0 1 3 6 10 10 45 50 66

Cuối cùng ta được biểu đồ:

0	0 1 3 6
1	0 0
2	
3	
4	5
5	0
6	6

Đơn vị tính của lá: 10 (triệu đồng)

Một số kỹ thuật như tách cành hay ghép lá có thể được sử dụng để việc mô tả được rõ rệt.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

- *Tách cành : Nếu số lá mỗi cành quá nhiều mà số cành ít*, ta có thể tách mỗi cành làm 2 cành : cành thấp (lá từ 0 đến 4) và cành cao (lá từ 5 đến 9). Ngoài ra cũng có thể tách mỗi cành làm 5 cành nhỏ: cành thứ nhất (lá 0 và 1), cành thứ hai (lá 2 và 3), cành thứ ba (lá 4 và 5), cành thứ tư (lá 6 và 7), cành thứ năm (lá 8 và 9).

Ví dụ, có biểu đồ cành và lá:

```
2 | 2 2 2 2 3 3 4 4 4 4 5 5 8 8 9 9
3 | 1 1 1 1 3 3 5 5 8 8 8 8 9
4 | 1 1 2 2 2 2 5 5 5 5 6 6 7 7 8
```

Ta có thể tách đôi cành như sau:

```
2 | 2 2 2 2 3 3 4 4 4 4
2 | 5 5 8 8 9 9
3 | 1 1 1 1 3 3
3 | 5 5 8 8 8 8 9
4 | 1 1 2 2 2 2
4 | 5 5 5 5 6 6 7 7 8
```

- *Ghép lá:* Nếu số lá quá nhiều trên mỗi cành, ta có thể ghép 2 lá giống nhau làm 1 lá đôi.

Ví dụ: Biểu đồ ở trên nếu không tách cành có thể ghép lá như sau:

```
2 | 2 2 3 4 4 5 8 9
3 | 1 1 3 5 8 8 9 &
4 | 1 2 2 5 5 6 7 8 &
```

Lá: lá đôi, ký hiệu lá chiếc &

2.3. Mô tả (tóm tắt) dữ liệu cho hai tiêu thức (biến) bằng bảng chéo và biểu đồ

2.3.1. Lập bảng phân phối kết hợp (bảng chéo) (crosstables)

Bảng chéo có dạng hình chữ nhật, trong đó các dòng trình bày các tổ (nhóm) của tiêu thức (biến) thứ nhất, các cột trình bày các tổ (nhóm) của tiêu thức (biến) thứ hai. Giao của các dòng và các cột là tần số (số quan sát) kết hợp của cả hai tiêu thức (biến).

Bảng chéo được dùng với 1 trong 3 trường hợp: Một tiêu thức (biến) là định tính và một tiêu thức (biến) là định lượng, cả hai tiêu thức (biến) là định tính, hoặc cả hai tiêu thức (biến) là định lượng. Việc xác định các tổ (nhóm) cho mỗi tiêu thức (biến) được tiến hành tương tự như đã trình bày trong mục 2.1.1 và mục 2.2.1. Tuy nhiên số lượng tổ (nhóm) theo từng tiêu thức (biến) được chọn sao cho tích của chúng (số tổ kết hợp) không quá nhiều. Có thể xác định số tổ kết hợp định hướng theo công thức gợi ý: $k = (2*n)^{0,333}$.

Ví dụ, có bảng phân phối chéo của hai tiêu thức mức độ ưa thích sản phẩm và độ tuổi của 360 người tiêu dùng như sau:

Mức độ ưa thích SP	Độ tuổi					Tổng cộng
	Dưới 16	16 - 25	26 - 45	46 - 60	Trên 60	
Không thích	15	32	18	25	5	95
Thích ít	5	8	20	38	8	79
Khá thích	2	7	30	42	15	96
Rất thích	1	5	10	50	24	90
Tổng cộng	23	52	78	155	52	360

Dựa vào bảng phân phối này, ta thấy rõ đặc điểm phân phối người tiêu dùng theo mức độ ưa thích ở từng độ tuổi và đặc điểm phân phối người tiêu dùng theo độ tuổi ở từng mức độ ưa thích. Qua đó, ta thấy được giữa hai tiêu thức này có biểu hiện của mối liên hệ nào đó chi phối hay không, nếu có thì mối liên hệ đó diễn ra theo chiều hướng nào.

Từ bảng chéo, ta có thể dễ dàng lập bảng phân phối riêng cho từng tiêu thức (biến). Chẳng hạn từ bảng chéo trên ta có thể lập bảng phân phối riêng cho tiêu thức (biến) mức độ ưa thích sản phẩm và tiêu thức (biến) độ tuổi. Ngoài ra, từ bảng chéo trên ta cũng có thể chuyển đổi dữ liệu trong bảng sang dạng tỉ lệ phần trăm theo dòng hay theo cột.Việc làm này giúp ta hiểu sâu hơn về mối quan hệ giữa hai tiêu thức (biến).

Bảng phân phối người tiêu dùng theo mức độ ưa thích sản phẩm

Mức độ ưa thích SP	Tần số	Tần suất	Tần suất (%)
Không thích	95	0,26	26
Thích ít	79	0,22	22
Khá thích	96	0,27	27
Rất thích	90	0,25	25
Tổng cộng	360	1,00	100

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Bảng phân phối người tiêu dùng theo độ tuổi

Độ tuổi	Tần số	Tần suất	Tần suất (%)
Dưới 16	23	0.06	6
16 - 25	52	0.14	14
26 - 45	78	0.22	22
46 - 60	155	0.43	43
Trên 60	52	0.14	14
Tổng cộng	360	1,00	100

Bảng chéo tỉ lệ phần trăm theo dòng

Mức độ ưa thích SP	Độ tuổi					Tổng cộng
	Dưới 16	16 - 25	26 - 45	46 - 60	Trên 60	
Không thích	15.8	33.7	18.9	26.3	5.3	100
Thích ít	6.3	10.1	25.3	48.1	10.1	100
Khá thích	2.1	7.3	31.3	43.8	15.6	100
Rất thích	1.1	5.6	11.1	55.6	26.7	100

Bảng chéo tỉ lệ phần trăm theo cột

Mức độ ưa thích SP	Độ tuổi				
	Dưới 16	16 - 25	26 - 45	46 - 60	Trên 60
Không thích	65.2	61.5	23.1	16.1	9.6
Thích ít	21.7	15.4	25.6	24.5	15.4
Khá thích	8.7	13.5	38.5	27.1	28.8
Rất thích	4.3	9.6	12.8	32.3	46.2
Tổng cộng	100	100	100	100	100

Ngoài dạng bảng chéo thông thường theo hai tiêu thức (biến) như trên, đôi khi, người ta còn lập bảng chéo cho ba hay bốn tiêu thức (biến). Khi đó, trên một cạnh của bảng chéo có thể có sự phân tổ (nhóm) kết hợp của hai tiêu thức (biến).

Ví dụ, có bảng phân phối chéo của ba tiêu thức là giới tính, mức độ ưa thích sản phẩm và độ tuổi của người tiêu dùng như sau:

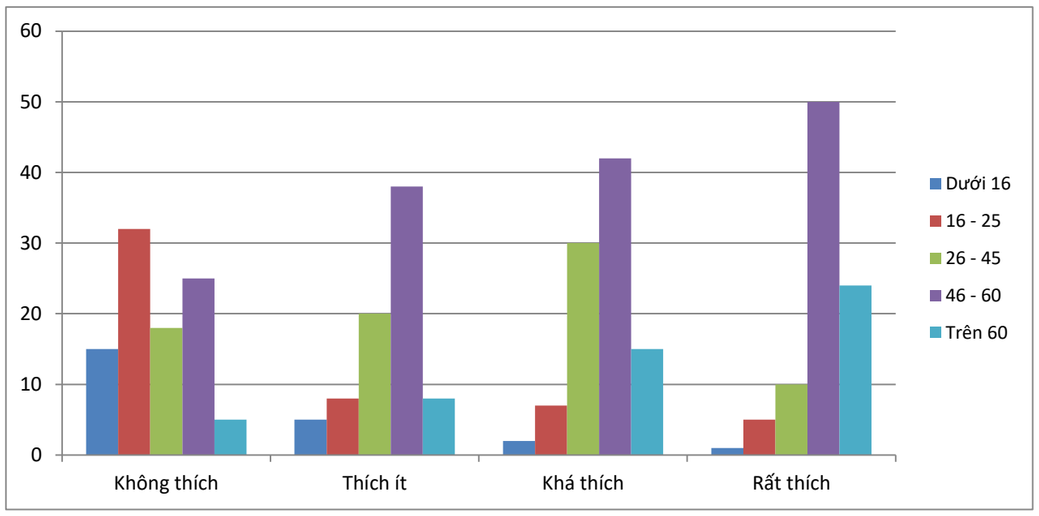
Giới tính và mức độ ưa thích SP	Độ tuổi					Tổng cộng
	Dưới 16	16 - 25	26 - 45	46 - 60	Trên 60	
Nam	23	52	78	155	52	360
Không thích	*15*	*32*	*18*	*25*	*5*	*95*
Thích ít	*5*	*8*	*20*	*38*	*8*	*79*
Khá thích	*2*	*7*	*30*	*42*	*15*	*96*
Rất thích	*1*	*5*	*10*	*50*	*24*	*90*
Nữ	14	44	136	92	36	322
Không thích	*11*	*22*	*28*	*12*	*6*	*95*
Thích ít	*2*	*10*	*32*	*28*	*5*	*79*
Khá thích	*1*	*8*	*40*	*32*	*11*	*96*
Rất thích	*0*	*4*	*36*	*30*	*14*	*90*
Tổng cộng	37	96	214	247	88	682

2.3.2. Trình bày bằng biểu đồ

Biểu đồ nhiều thanh cạnh nhau (Side by side bar charts)

Dữ liệu trên bảng phân phối chéo theo tần số của hai tiêu thức (không bao gồm dòng tổng cộng và cột tổng cộng) được dùng để vẽ biểu đồ nhiều thanh cạnh nhau. Trong đó, từng nhóm thanh nhiều màu thể hiện một dòng dữ liệu trong bảng.

Ví dụ, bảng phân phối chéo của hai tiêu thức mức độ ưa thích sản phẩm và độ tuổi của 360 người tiêu dùng ở trên có thể được trình bày trên biểu đồ nhiều thanh cạnh nhau như sau:



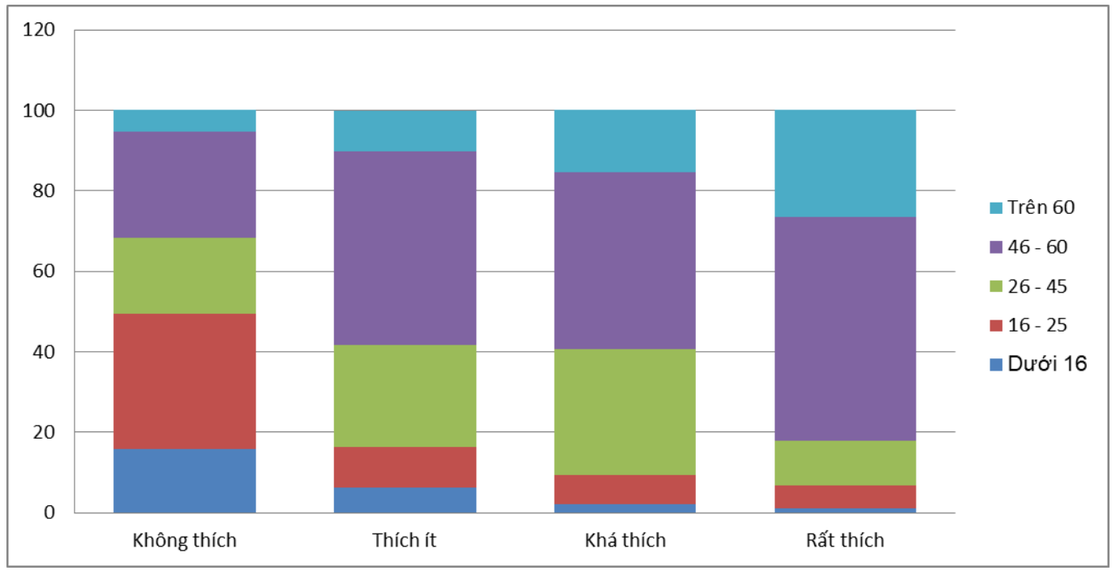
Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

So sánh chiều cao giữa các thanh có màu khác nhau trong cùng một nhóm thanh và các thanh cùng màu giữa các nhóm thanh giúp ta nhận thức được đặc điểm bên trong mỗi tiêu thức (biến) và mối liên hệ giữa hai tiêu thức (biến).

Biểu đồ nhiều thanh chồng nhau (Stacked bar charts)

Dữ liệu trên bảng chéo tỉ lệ phần trăm theo dòng (không bao gồm cột tổng cộng) được dùng để vẽ biểu đồ nhiều thanh chồng nhau. Trong đó mỗi thanh nhiều màu thể hiện một dòng dữ liệu phần trăm trong bảng.

Ví dụ, bảng phân phối chéo tỉ lệ phần trăm theo dòng của 360 người tiêu dùng ở trên có thể được trình bày trên biểu đồ nhiều thanh cạnh nhau như sau:



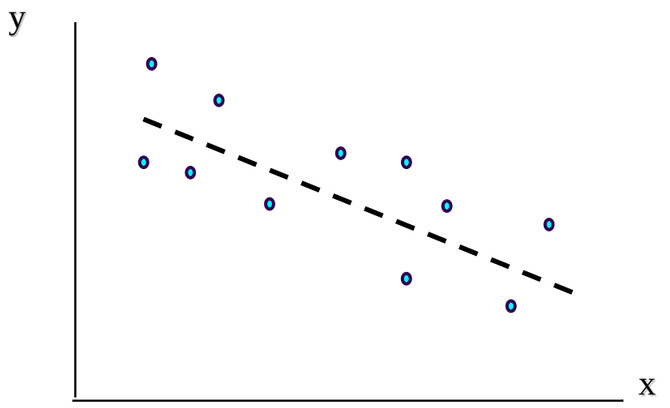
So sánh chiều cao giữa các đoạn thanh có màu khác nhau trong cùng một thanh và so sánh các đoạn thanh cùng màu giữa các thanh giúp ta nhận thức được đặc điểm bên trong mỗi biến và mối liên hệ giữa hai biến.

Đồ thị phân tán và đường xu hướng (scatter plots and trendline)

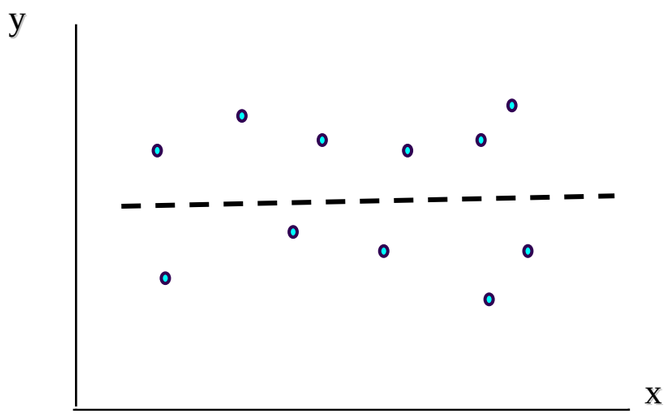
Đồ thị phân tán thường được dùng để mô tả phân phối kết hợp giữa hai tiêu thức (biến) định lượng hoặc của một tiêu thức (biến) định lượng và một tiêu thức (biến) định tính trên thang đo thứ bậc. Đồ thị này được vẽ dựa vào tập hợp dữ liệu gốc.

Một tiêu thức (biến) được trình bày ở trục hoành (x) và tiêu thức biến còn lại trên trục tung (y). Mỗi chấm trên đồ thị là một quan sát chung giữa 2 biến.

Đường xu hướng là một đường thể hiện xu hướng phân phối của các điểm trên đồ thị, cung cấp một xấp xỉ về mối quan hệ giữa hai biến được trình bày.

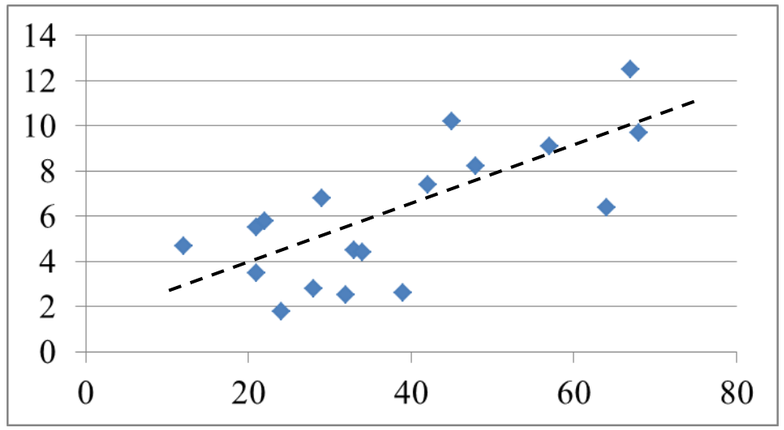


Có liên hệ nghịch giữa hai biến



Không có liên hệ giữa hai hiến

Ví dụ, đồ thị phân tán thể hiện mối liên hệ giữa thời gian gặp sự cố (x) và giá trị sản phẩm (y) trong dữ liệu về 18 đơn thư khiếu nại ở trang 3 được trình bày như sau:



Có liên hệ thuận rõ rệt giữa hai biến

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

2.4. Mô tả một tiêu thức (biến) định lượng bằng các chỉ tiêu thống kê

Một khối lượng lớn dữ liệu của tiêu thức (biến) định lượng có thể được mô tả tóm tắt thông qua vài chỉ tiêu thống kê biểu hiện những nét đặc trưng nhất của chúng. Đây là phương pháp mô tả dữ liệu định lượng cô đọng nhất, khái quát nhất. Nó giúp ta lược bỏ những thứ rườm rà, khó hiểu, nhờ đó nhận thức được những đặc trưng cốt lõi nhất của hiện tượng.

2.4.1. Các chỉ tiêu mô tả vị trí trung tâm (khuynh hướng hội tụ)

Khi cần nhận thức khái quát nhất về một tiêu thức (biến) định lượng điều trước tiên chúng ta cần biết là tâm điểm của biến. Các chỉ tiêu thống kê vị trí trung tâm (khuynh hướng hội tụ) đáp ứng yêu cầu này. Có ba loại chỉ tiêu thường dùng nhất là Số trung bình, Trung vị và Mốt.

a. Số trung bình (Mean): Được xác định bằng cách lấy tổng các lượng biến (quan sát) của tiêu thức chia đều cho số đơn vị tổng thể.

- Trường hợp dữ liệu mẫu:

+ *Với dữ liệu không có tần số*: $\bar{x} = \dfrac{\sum x_i}{n}$

Trong đó: $\bar{x}$: Số trung bình mẫu (đọc là x ngang)
x_i : Lượng biến (các quan sát)
n : Số đơn vị tổng thể (số quan sát)

Ví dụ: Với dữ liệu về 18 đơn thư khiếu nại ở trang 3, tuổi trung bình của các khách hàng:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{22 + 26 + 25 + ... + 27}{18} = 28.89 \quad \text{(tuổi)}$$

+ *Với dữ* liệu có tần số*:*

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

Trong đó: x_i : Lượng biến (các giá trị có tần số)
f_i : Tần số (hoặc tần suất)

Ví dụ, có bảng phân phối tần số về độ tuổi của các khách hàng trong dữ liệu 18 đơn thư khiếu nại ở trang 3 như sau:

Độ tuổi	Tần số (f_i)
22	1
25	4
26	7
27	4
48	1
60	1

Độ tuổi trung bình của các khách hàng:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{22\times1 + 25\times4 + \cdots + 60\times1}{1 + 4 + \cdots + 1} = 28{,}89 \text{ (tuổi)}$$

+ *Với dữ liệu được phân tổ có khoảng cách tổ:*

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

Trong đó: f_i : Tần số (hoặc tần suất)
$x_i = (x_{imin} + x_{imax})/2$: Trị số giữa tổ i

Ví dụ, có bảng phân phối tần số 600 người tiêu dùng:

Độ tuổi	Tần số (f_i)	Trị số giữa (x_i)
Dưới 16	20	10,5
16 – 25	180	20,5
26 – 45	260	35,5
46 – 60	100	53,0
Trên 60	40	68,0

Tuổi trung bình của 600 người tiêu dùng:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{10{,}5 \cdot 20 + 20{,}5 \cdot 180 + 35{,}5 \cdot 260 + 53{,}0 \cdot 100 + 68{,}0 \cdot 40}{20 + 180 + 260 + 100 + 40}$$

$$= 35{,}25 \text{ (tuổi)}$$

Trong đó, tổ đầu và tổ cuối gọi là tổ mở có trị số giữa được tính dựa vào khoảng cách tổ của tổ gần chúng nhất như sau:

Trị số giữa của tổ thứ nhất: (15 + (15-9)) / 2 = 10,5
Trị số giữa của tổ cuối: (61 + (61+14)) / 2 = 68,0

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

- Trường hợp dữ liệu tổng thể: Các công thức tính số trung bình tổng thể tương tự số trung bình mẫu. Tuy nhiên để phân biệt người ta thường dùng ký hiệu khác như sau.

+ *Với dữ liệu không có tần số*: $\mu = \dfrac{\sum x_i}{N}$

Trong đó: μ : Số trung bình tổng thể
x_i : Lượng biến (quan sát)
N : Số đơn vị tổng thể của tổng thể

+ *Với dữ* liệu có tần số*:* $\mu = \dfrac{\sum x_i f_i}{\sum f_i}$

Trong đó: x_i : Lượng biến (các giá trị có tần số)
f_i : Tần số (hoặc tần suất)

Số trung bình là chỉ tiêu mô tả vị trí trung tâm tốt nhất và thường được dùng nhất. Tuy nhiên, chỉ tiêu này chịu ảnh hưởng mạnh của các lượng biến (quan sát) cực biên. Do đó, khi dữ liệu có các lượng biến (quan sát) cực biên thì tính chất định tâm của số đo này không được cao. Khi đó, nó cần được bổ sung thêm Trung vị hay Mốt.

Chú ý: Một số đo khác có ý nghĩa tương tự Số trung bình mô tả ở trên là *Số trung bình có trọng số. Số đo này thường được dùng cho các mục đích quản lý kinh tế khác đòi hỏi độ chính xác cao.*

Số trung bình có trọng số: $\bar{x} = \dfrac{\sum x_i w_i}{\sum w_i}$

Trong đó: x_i *: Lượng biến (quan sát)*
w_i *: Trọng số (quyền số)*

Trọng số là đại lượng thể hiện tầm quan trọng của từng lượng biến (quan sát). Tùy mục đích cụ thể, trọng số được chọn một cách thích hợp.

Trong lĩnh vực kinh tế, Số trung bình có trọng số thường hay được dùng để phản ảnh trị số trung tâm của các tiêu thức có thể viết dưới dạng $x_i = \dfrac{M_i}{w_i}$, chẳng hạn như lợi suất trên vốn, lợi suất trên doanh số, tốc độ chu chuyển của vốn lưu động, đơn giá, giá thành đơn vị sản phẩm ...

Ví dụ, có dữ liệu như sau về các đại lý của một doanh nhiệp:

Đại lý	Tỉ trọng vốn lưu động trong vốn kinh doanh (%)	Vốn kinh doanh (tỉ đồng)
1	20	1,0
2	15	0,6
3	28	0,4
4	30	0,8
5	16	0,7

Tỉ trọng vốn lưu động trung bình có trọng số:

$$\bar{x} = \frac{\sum x_i w_i}{\sum w_i} = \frac{20*1{,}0 + 15*0{,}6 + 28*0{,}4 + 30*0{,}8 + 16*0{,}7}{1{,}0 + 0{,}6 + 0{,}4 + 0{,}8 + 0{,}7}$$

$= 21{,}54$ (%)

Trong đó: x_i : Tỉ trọng vốn lưu động từng đại lý
w_i : Trọng số (vốn kinh doanh từng đại lý)

Lưu ý:

1) Số trung bình cộng có trọng số có thể dùng trọng số là tỉ trọng mỗi bộ phận trong tổng thể:

$$\bar{x} = \frac{\sum x_i w_i}{\sum w_i} = \frac{\sum x_i d_i}{\sum d_i} \qquad \text{với} \quad d_i = \frac{w_i}{\sum w_i}$$

Ví dụ, có dữ liệu như sau về các đại lý của một doanh nhiệp:

Đại lý	Lợi suất tính trên doanh số (%)	Tỉ trọng doanh số (%)
1	8	10
2	10	40
3	6	15
4	9	5
5	7	30

Lợi suất trung bình có trọng số:

$$\bar{x} = \frac{\sum x_i w_i}{\sum w_i} = \frac{8*10 + 10*40 + 6*15 + 9*5 + 7*30}{10 + 40 + 15 + 5 + 30} = 8{,}25 \ (\%)$$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Trong đó: x_i : Lợi suất từng đại lý

w_i : Trọng số (tỉ trọng doanh số từng đại lý)

2) Số trung bình có trọng số ít được dùng làm số đo vị trí trung tâm trong mô tả tóm tắt dữ liệu chéo vì gặp khó khăn trong suy rộng dữ liệu mẫu cho tổng thể và không tương thích với cách tính các số đo định tâm khác như Trung vị hay Mốt.

b. Trung vị (Median): *Là lượng biến (quan sát) đứng ở vị trí giữa của dãy số lượng biến (quan sát) đã được sắp xếp, chia dãy số lượng biến (quan sát) thành hai phần bằng nhau.*

Nếu n lẽ: $M_e = x_{(\frac{n+1}{2})}$

Nếu n chẵn: $M_e = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$

Ví dụ 1: Với dữ liệu về 18 đơn thư khiếu nại ở trang 3, giá trị sản phẩm của các khách hàng khiếu nại được sắp xếp theo độ lớn như sau:

1,8	2,5	2,6	2,8	3,5	4,4	4,5	4,7	5,5
5,8	6,4	6,8	7,4	8,2	9,1	9,7	10,2	12,5

Giá trị sản phẩm trung vị:

$$M_e = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{x_9 + x_{10}}{2} = \frac{5,5+5,8}{2} = 5,65 \ (tr)$$

Ví dụ 2, có bảng phân phối 500 người tiêu dùng theo số thành viên trong gia đình.

Số thành viên trong gia đình	Tần số (f_i)	Tần số tích luỹ (S_i)
1	10	10
2	80	90
3	200	290
4	120	410
5	80	490
6 trở lên	10	500

$$M_e = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{x_{250} + x_{251}}{2} = \frac{3+3}{2} = 3 \text{ (thành viên)}$$

c. Mốt (Mode): *Là lượng biến (quan sát) xuất hiện nhiều nhất trong tập hợp dữ liệu.*

Với dữ liệu trong bảng phân phối tần số, mốt là lượng biến có tần số lớn nhất.

Ví dụ, xét bảng phân phối 500 người tiêu dùng theo số thành viên trong gia đình kế trên.

$M_o = 3$ (thành viên)

Chú ý: Mốt có ý nghĩa định tâm rất kém với dữ liệu không có điểm tập trung hay có nhiều điểm tập trung. Không nên dùng Mốt trong những trường hợp này.

2.4.2. Các chỉ tiêu mô tả độ phân tán

Các chỉ tiêu mô tả vị trí trung tâm chỉ phản ảnh được tâm điểm phân phối của một tập hợp dữ liệu. *Một đặc trưng phân phối khác không thể bỏ qua mà phải được mô tả tóm tắt đó là dữ liệu phân tán như thế nào quanh tâm điểm của nó.* Có bốn loại chỉ tiêu thống kê độ phân tán thường dùng nhất là khoảng biến thiên, độ lệch tuyệt đối trung bình, phương sai và độ lệch chuẩn. Ngoài ra, hệ số biến thiên được dùng để phản ảnh độ phân tán tương đối.

a. Khoảng biến thiên (Range):

$$R = x_{max} - x_{min}$$

Trong đó: x_{max} : Lượng biến lớn nhất

x_{min} : Lượng biến nhỏ nhất

Chỉ tiêu thống kê này phản ảnh độ lớn khoảng phân tán của dữ liệu. Tuy nhiên, nó không xét đến tất cả các lượng biến nên kém chặt chẽ.

Ví dụ : Với dữ liệu về 18 đơn thư khiếu nại ở trang 3, khoảng biến thiên giá trị sản phẩm của các khách hàng khiếu nại:

$$R = x_{max} - x_{min} = 12,5 - 1,8 = 10,7$$

b. Độ lệch tuyệt đối trung bình (Mean absolute deviation)

- *Với dữ liệu không có tần số*: $\bar{d} = \frac{\sum |x_i - \bar{x}|}{n}$ với $\bar{x} = \frac{\sum x_i}{n}$

Trong đó: x_i : Lượng biến

n : Số đơn vị tổng thể

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Ví dụ: Với dữ liệu về 18 đơn thư khiếu nại ở trang 3, độ lệch tuyệt đối trung bình về tuổi của các khách hàng:

$$\bar{d} = \frac{\sum|x_i - \bar{x}|}{n} = \frac{|22-28{,}9| + |25-28{,}9| + ... + |60-28{,}9|}{18} = 5{,}6$$

Trong đó: $\bar{x} = \frac{\sum x_i}{n} = \frac{22+26+25+...+27}{18} = 28{,}9$

- *Với dữ liệu có tần số*:

$$\bar{d} = \frac{\sum|x_i - \bar{x}| f_i}{\sum f_i} \quad \text{với} \quad \bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

Trong đó: x_i : Lượng biến (các trị số rút gọn hay trị số giữa)
f_i : Tần số

Chỉ tiêu này đo độ phân tán tuyệt đối trung bình quanh số trung bình. Tuy nhiên, nhược điểm của nó là chứa dấu tuyệt đối nên gặp khó khăn khi đưa vào các công thức tính toán khác.

b. Phương sai (Variance):

- ***Phương sai mẫu:***

+ Với dữ liệu không có tần số: $s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$

Trong đó: x_i : Lượng biến (các quan sát)
n : Số đơn vị tổng thể (số quan sát) của mẫu

Ví dụ : Với dữ liệu về 18 đơn thư khiếu nại ở trang 3, phương sai giá trị sản phẩm của các khách hàng khiếu nại:

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{(2{,}5-6{,}02)^2 + (1{,}8-6{,}02)^2 + ... + (2{,}6-6{,}02)^2}{18-1} = 9{,}15$$

Trong đó: $\bar{x} = \frac{\sum x_i}{n} = \frac{2{,}5+1{,}8+12{,}5+...+2{,}6}{18} = 6{,}02$

+ Với dữ liệu có tần số:

$$s^2 = \frac{\sum(x_i - \bar{x})^2 f_i}{n-1}$$

Trong đó: x_i : Lượng biến (các trị số có tần số hay trị số giữa)
f_i : Tần số

Ví dụ, có bảng phân phối chi tiêu của một mẫu 500 người tiêu dùng:

Chi tiêu (tr)	Tần số (f_i)
10,5	50
20,5	80
35,5	120
53,0	170
68,0	80

Phương sai chi tiêu của người tiêu dùng:

$$s^2 = \frac{\sum(x_i - \bar{x})^2 f_i}{n-1}$$

$$= \frac{(10{,}5-41{,}8)^2 * 50 + (20{,}5-41{,}8)^2 * 80 + ... + (68{,}0-41{,}8)^2 * 80}{500-1} = 333$$

Trong đó: $\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{10{,}5.50 + 20{,}5.80 + ... + 68{,}0.80}{50+80+...+80} = 41{,}8$

- Phương sai tổng thể: $\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$

Trong đó: x_i : Lượng biến
N : Số đơn vị tổng thể của tổng thể

Phương sai phản ánh độ phân tán của mọi lượng biến quanh Số trung bình. Đây là số đo độ phân tán rất tốt, rất hay được dùng. Tuy nhiên, đơn vị tính của nó là bình phương đơn vị tính của tiêu thức ban đầu nên có tính trừu tượng cao, khó nhận thức.

d. Độ lệch chuẩn (Standard deviation):

Độ lệch chuẩn tổng thể:

$$\sigma = \sqrt{\sigma^2}$$

Độ lệch chuẩn mẫu:

$$s = \sqrt{s^2}$$

Ví dụ, độ lệch chuẩn về chi tiêu của 500 người tiêu dùng ở ví dụ trên:

$$s = \sqrt{s^2} = \sqrt{333} = 18{,}2$$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Đây là chỉ tiêu đo độ phân tán tốt nhất, thường được dùng nhất cho một tập hợp dữ liệu của một tiêu thức.

Chú ý:

(1) Qui tắc Tchebychev: Với một tổng thể bất kỳ, số quan sát có giá trị nằm quanh số trung bình với khoảng cách $\pm m.\sigma$ với mọi m>1 chiếm ít nhất là (1-1/m^2).

Khoảng cách	$\pm 1{,}5.\sigma$	$\pm 2{,}0.\sigma$	$\pm 2{,}5.\sigma$	$\pm 3{,}0.\sigma$
Tỉ lệ số quan sát ít nhất (%)	55,6	75,0	84,0	88,9

(2) Qui tắc phát hiện các giá trị bất thường: Giá trị bất thường (hay giá trị cực biên) là những giá trị lớn hay nhỏ một cách bất bình thường. Dựa trên qui tắc Tchebychev, những quan sát được xem là bất thường khi có giá trị: $x > \bar{x} + 3s$ hoặc $x < \bar{x} - 3s$.

(3) Xử lý giá trị bất thường: Xem xét kỹ lưỡng các giá trị được phát hiện là bất thường để có biện pháp xử lý phù hợp.

- Nếu có sai sót trong khâu thu thập dữ liệu: Hiệu chỉnh hay thu thập lại.
- Nếu giá trị bất thường không thuộc đối tượng nghiên cứu: Loại bỏ.
- Nếu giá trị bất thường thuộc đối tượng nghiên cứu và thu thập đúng: Giữ lại.

d. Hệ số biến thiên (Coefficient of variation):

Hệ số biến thiên tổng thể: $C_V = \frac{\sigma}{\mu}.100$ (%)

Hệ số biến thiên mẫu: $C_V = \frac{s}{\bar{x}}.100$ (%)

Ví dụ, với dữ liệu mẫu 500 người tiêu dùng ở ví dụ trên:

$$C_V = \frac{18{,}2}{41{,}8}.100 = 43{,}7\ (\%)$$

Đây là chỉ tiêu đo độ phân tán tương đối. Nó thường được dùng để so sánh độ phân tán của cùng một tiêu thức trên các tập dữ liệu khác nhau (về địa điểm hay thời gian) hoặc của các tiêu thức khác nhau.

2.4.3. Tứ phân vị, biểu đồ hộp

Tứ phân vị bao gồm ba số đo chia dãy lượng biến (đã sắp xếp) thành bốn phần bằng nhau.

+ Tứ phân vị thứ nhất: $Q_1 = x_{(n+1)/4} = x_{(n_1+t_1/4)} = x_{n_1} + \frac{t_1}{4}(x_{(n_1+1)} - x_{n_1})$

+ Tứ phân vị thứ hai: $Q_2 = x_{2(n+1)/4} = x_{(n_2+t_2/4)} = x_{n_2} + \frac{t_2}{4}(x_{(n_2+1)} - x_{n_2})$

+ Tứ phân vị thứ ba: $Q_3 = x_{3(n+1)/4} = x_{(n_3+t_3/4)} = x_{n_3} + \frac{t_3}{4}(x_{(n_3+1)} - x_{n_3})$

Trong đó, n_1, n_2, n_3 là các số nguyên dương, t_1, t_2, t_3 là các phần dư (0, 1, 2 hoặc 3).

Ví dụ: Với dữ liệu về 18 đơn thư khiếu nại ở trang 3, giá trị sản phẩm của các khách hàng khiếu nại được sắp xếp theo độ lớn như sau:

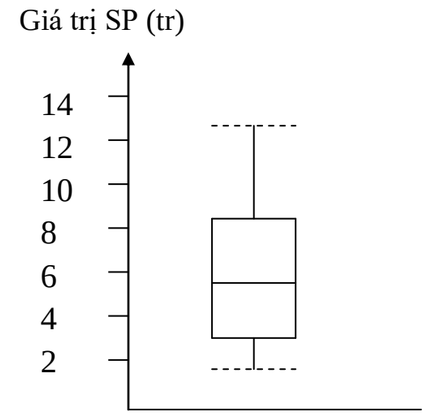
1,8	2,5	2,6	2,8	3,5	4,4	4,5	4,7	5,5
5,8	6,4	6,8	7,4	8,2	9,1	9,7	10,2	12,5

$Q_1 = x_{(n+1)/4} = x_{19/4} = x_4 + 3/4(x_5 - x_4) = 2{,}8 + ¾(3{,}5\text{-}2{,}8) = 3{,}3$

$Q_2 = x_{2(n+1)/4} = x_{38/4} = x_9 + 2/4(x_{10} - x_9) = 5{,}5 + 2/4(5{,}8\text{-}5{,}5) = 5{,}7$

$Q_3 = x_{3(n+1)/4} = x_{57/4} = x_{14} + 1/4\,(x_{15} - x_{14}) = 8{,}2 + ¼(9{,}1\text{-}8{,}2) = 8{,}4$

Dựa vào các tứ phân vị, ta có thể mô tả tóm tắt dữ liệu dưới dạng biểu đồ hộp như sau:



Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Trong đó, đáy của hộp chữ nhật chỉ Tứ phân vị thứ nhất. Gạch ngang ở giữa hộp chữ nhật chỉ Trung vị. Cạnh trên của hộp chữ nhật chỉ Tứ phân vị thứ ba. Hai đường gạch đứt ở trên cùng và dưới cùng (ria) chỉ lượng biến lớn nhất và lượng biến nhỏ nhất.

Trường hợp dữ liệu có các lượng biến cực biên bất thường thì những lượng biến này được mô tả cụ thể bằng các điểm riêng biệt ở phía trên đường gạch đứt trên (ria trên) và phía dưới đường gạch đứt phía dưới (ria dưới). Lúc này, lượng biến lớn nhất và lượng biến nhỏ nhất được xác định dựa trên những lượng biến còn lại.

Chú ý: Qui tắc phát hiện giá trị bất thường: $x > \bar{x} + 3s$ hoặc $x < \bar{x} - 3s$. Qui tắc tương đương: $x > Q_3 + 1{,}5\,(Q_3 - Q_1)$ hoặc $x < Q_1 - 1{,}5\,(Q_3 - Q_1)$.

Ở ví dụ trên: $Q_3 + 1{,}5\,(Q_3 - Q_1) = 8{,}4 + 1{,}5(8{,}4 - 3{,}3 = 16{,}05$

$Q_1 - 1{,}5\,(Q_3 - Q_1) = 3{,}3 - 1{,}5(8{,}4 - 3{,}3) = -4{,}35.$

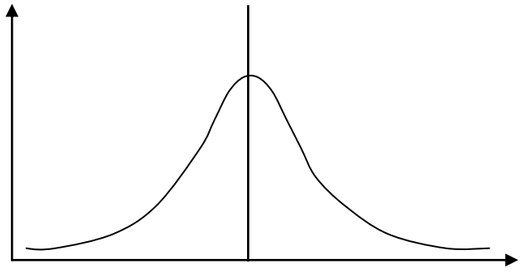
Do đó không có giá trị bất thường.

Biểu đồ hộp cho ta nhận thức có tính hình ảnh về đặc điểm phân phối của hiện tượng là phân tán hay tập trung ở đâu.

2.5. Các chỉ tiêu mô tả hình dáng phân phối một tiêu thức định lượng:

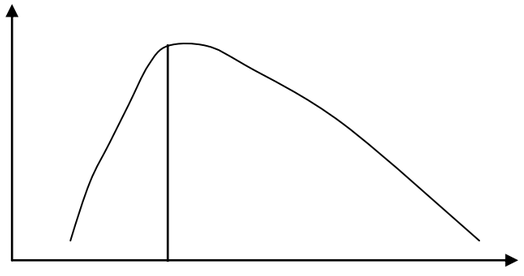
Hình dáng phân phối là một trong những đặc trưng quan trọng của dãy số phân phối. *Nó cho ta hình ảnh phân phối của tổng thể trên một tiêu thức định lượng.* Hình dáng phân phối thể hiện rõ qua tính đối xứng và độ dốc của phân phối.

- Phân phối đối xứng:



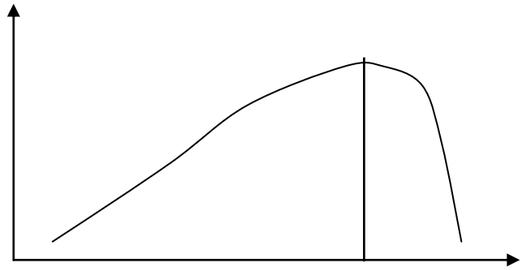
$\bar{x} = M_e = M_o$

- Phân phối lệch phải:



$M_o < M_e < \bar{x}$

- Phân phối lệch trái:



$\bar{x} < M_e < M_o$

- Để mô tả tính đối xứng của phân phối, người ta thường sử dụng hệ số SKEWNESS:

$$SKEW = \frac{\sqrt{n(n-1)}}{n-2} H$$

Trong đó:

$$H = \frac{\frac{\sum(x_i - \bar{x})^3}{n}}{\left(\frac{\sum(x_i - \bar{x})^2}{n}\right)^{\frac{3}{2}}}$$

Hệ số SKEW bằng không, phân phối đối xứng. Hệ số SKEW dương càng lớn, phân phối lệch phải càng nhiều. Hệ số SKEW âm càng lớn phân phối lệch trái càng nhiều.

- Để mô tả độ nhọn của phân phối, người ta thường sử dụng hệ số KURTOSIS:

$$KURT = \frac{(n-1)(n+1)}{(n-2)(n-3)} K - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Trong đó:

$$K = \frac{\frac{\sum(x_i - \bar{x})^4}{n}}{\left(\frac{\sum(x_i - \bar{x})^2}{n}\right)^2}$$

Hệ số KURT bằng không, phân phối có độ dốc như phân phối chuẩn. Hệ số KURT dương càng lớn, phân phối càng dốc hơn phân phối chuẩn. Hệ số KURT âm càng lớn phân phối càng ít dốc hơn phân phối chuẩn.

Ví dụ: Xét dữ liệu về giá trị sản phẩm của 18 đơn thư khiếu nại ở trang 3.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Khách hàng	Gía trị sphẩm	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
1	2,5	-3,52	12,41	-43,70	153,91
2	1,8	-4,22	17,83	-75,27	317,81
3	12,5	6,48	41,96	271,82	1760,78
4	4,5	-1,52	2,32	-3,53	5,37
5	2,8	-3,22	10,38	-33,46	107,80
6	6,4	0,38	0,14	0,05	0,02
7	10,2	4,18	17,45	72,92	304,64
8	3,5	-2,52	6,36	-16,05	40,47
9	6,8	0,78	0,60	0,47	0,37
10	5,5	-0,52	0,27	-0,14	0,07
11	4,7	-1,32	1,75	-2,31	3,06
12	8,2	2,18	4,74	10,33	22,49
13	9,1	3,08	9,47	29,15	89,73
14	7,4	1,38	1,90	2,62	3,60
15	5,8	-0,22	0,05	-0,01	0,00
16	4,4	-1,62	2,63	-4,27	6,93
17	9,7	3,68	13,53	49,75	182,95
18	2,6	-3,42	11,71	-40,08	137,16
Cộng	108,4	0	155,51	218,30	3137,16
Tr. bình	6,02	0,00	8,64	12,13	174,29

Với kết quả tính được ở bảng trên, ta có:

$$H = \frac{\frac{\sum(x_i - \bar{x})^3}{n}}{\left(\frac{\sum(x_i - \bar{x})^2}{n}\right)^{\frac{3}{2}}} = \frac{12,13}{8,64^{\frac{3}{2}}} = 0,48$$

$$SKEW = \frac{\sqrt{n(n-1)}}{n-2} H = \frac{\sqrt{18(18-1)}}{18-2} 0,48 = 0,52$$

SKEW dương, và khá gần 0. Do đó, phân phối giá trị sản phẩm có phần hơi lệch phải.

$$K = \frac{\frac{\sum(x_i - \bar{x})^4}{n}}{\left(\frac{\sum(x_i - \bar{x})^2}{n}\right)^2} = \frac{174,29}{8,64^2} = 2,33$$

$$KURT = \frac{(n-1)(n+1)}{(n-2)(n-3)} K - \frac{3(n-1)^2}{(n-2)(n-3)}$$

$$= \frac{(18-1)(18+1)}{(18-2)(18-3)} 2,33 - \frac{3(18-1)^2}{(18-2)(18-3)} = -0,48$$

Hệ số KURT âm và khá gần 0 nên phân phối giá trị sản phẩm có phần ít dốc hơn phân phối chuẩn một ít.

Kết hợp hai đặc điểm mô tả trên cho thấy hình dáng phân phối giá trị sản phẩm là gần xấp xỉ chuẩn.

2.6. Các chỉ tiêu mô tả mối liên hệ tương quan giữa hai tiêu thức định lượng

Giữa hai tiêu thức định lượng trong cùng một tổng thể nghiên cứu có thể có mối liên hệ tuyến tính ở một mức độ và chiều hướng nào đó. Để đo lường mức độ và chiều hướng mối liên hệ tương quan tuyến tính này, người ta thường dùng Hiệp phương sai, Hệ số tương quan Pearson, Hệ số tương quan hạng Spearman.

2.6.1. Hiệp phương sai: Hiệp phương sai giữa hai tiêu thức định lượng x, y được xác định theo công thức:

- Hiệp phương sai tổng thể: $\sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N}$

- Hiệp phương sai mẫu: $S_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$

Ví dụ: Xét mối liên hệ tương quan giữa tuổi và giá trị sản phẩm của khách hàng khiếu nại ở trang 3, ta có:

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Đơn thư	Tuổi (x)	G.trị sp (y)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	22	2,5	-6,89	-3,52	24,26	47,46	12,41
2	26	1,8	-2,89	-4,22	12,20	8,35	17,83
3	25	12,5	-3,89	6,48	-25,19	15,12	41,96
4	27	4,5	-1,89	-1,52	2,88	3,57	2,32
5	26	2,8	-2,89	-3,22	9,31	8,35	10,38
6	26	6,4	-2,89	0,38	-1,09	8,35	0,14
7	25	10,2	-3,89	4,18	-16,25	15,12	17,45
8	27	3,5	-1,89	-2,52	4,76	3,57	6,36
9	26	6,8	-2,89	0,78	-2,25	8,35	0,60
10	48	5,5	19,11	-0,52	-9,98	365,23	0,27
11	26	4,7	-2,89	-1,32	3,82	8,35	1,75
12	25	8,2	-3,89	2,18	-8,47	15,12	4,74
13	26	9,1	-2,89	3,08	-8,89	8,35	9,47
14	25	7,4	-3,89	1,38	-5,36	15,12	1,90
15	60	5,8	31,11	-0,22	-6,91	967,90	0,05
16	27	4,4	-1,89	-1,62	3,06	3,57	2,63
17	26	9,7	-2,89	3,68	-10,62	8,35	13,53
18	27	2,6	-1,89	-3,42	6,46	3,57	11,71
Tổng	520,0	108,4	0,00	0,00	-28,26	1513,78	155,51
TB	28,89	6,02	0,00	0,00	-1,57	84,10	8,64

Ý nghĩa:

+ $S_{xy} = 0$ giữa hai tiêu thức x, y không có liên hệ tương quan.

+ $S_{xy} > 0$ giữa hai tiêu thức x, y có liên hệ tương quan thuận chiều.

+ $S_{xy} < 0$ giữa hai tiêu thức x, y có liên hệ tương quan nghịch chiều.

Ví dụ. Dựa vào kết quả tính được trong bảng trên, ta có:

$$S_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-28,26}{18-1} = -1,66$$

Vậy giữa tuổi và giá trị sản phẩm mà khách hàng khiếu nại có liên hệ tương quan nghịch chiều.

2.6.2. Hệ số tương quan Pearson: Hệ số này còn được gọi là hệ số tương quan đơn hay hệ số tương quan riêng bậc không. Công thức tính hệ số tương quan Pearson giữa tiêu thức x và tiêu thức y như sau:

- Hệ số tương quan tổng thể: $\rho = \frac{\sigma_{xy}}{\sigma_x . \sigma_y}$

- Hệ số tương quan mẫu: $r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{S_{xy}}{S_x S_y}$

Tính chất :

* $-1 \le r \le +1$
* $r = 0 \iff S_{xy} = 0$
* r không phản ảnh được tương quan phi tuyến
* r không phản ảnh quan hệ *độc lập-phụ thuộc* hay quan hệ *nhân-quả*

Ý nghĩa:

* r càng gần 1: Tương quan thuận càng mạnh.
* r càng gần -1: Tương quan nghịch càng mạnh.
* r càng gần 0: Tương quan càng yếu.

Ví dụ: Hệ số tương quan Pearson giữa tuổi và giá trị sản phẩm của khách hàng khiếu nại ở bảng trang 3:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{-28,26}{\sqrt{1513,78 * 155,51}} = -0,058$$

Vậy giữa tuổi và giá trị sản phẩm của khách hàng khiếu nại có liên hệ tương quan nghịch nhưng không đáng kể.

2.6.3. Hệ số tương quan hạng Spearman: Hệ số tương quan hạng Spearman được xây dựng dựa trên cơ sở xếp hạng độ lớn của các lượng biến của từng tiêu thức chứ không dựa trực tiếp vào các lượng biến này. Do đó, để tính hệ số tương quan hạng Spearman cần lập bảng xếp hạng. *Trong đó, hạng được xếp riêng cho từng tiêu thức theo thứ tự tăng dần của các lượng biến. Những giá trị bằng nhau sẽ nhận hạng trung bình của chúng.*

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Quan sát	Tiêu thức 1 (X)	Tiêu thức 2 (Y)	Hạng tiêu thức 1	Hạng tiêu thức 2	Chênh lệch hạng (X-Y)	d^2_i
1	x_1	y_1			d_1	d^2_1
2	x_2	y_2			d_2	d^2_2
3	x_3	y_3			d_3	d^2_3
...	...	...			...	...
n	x_n	y_n			d_n	d^2_n

Hệ số tương quan hạng Spearman: $$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

Tính chất :

* $-1 \le r \le +1$
* r không phản ảnh được liên hệ phi tuyến
* r không phản ảnh quan hệ *độc lập-phụ thuộc* hay quan hệ *nhân-quả*

* So với hệ số tương quan Pearson, hệ số tương quan hạng Spearman có phần yếu hơn. Tuy nhiên nó có thể sử dụng cho dữ liệu không cần phân phối chuẩn kết hợp của hai tiêu thức.

Ý nghĩa:

* r càng gần 1: Tương quan thuận càng mạnh.
* r càng gần -1: Tương quan nghịch càng mạnh.
* r càng gần 0: Tương quan càng yếu.
* r = 0 : Không có liên hệ tương quan

Ví dụ: Hệ số tương quan hạng giữa tuổi và giá trị sản phẩm của khách hàng khiếu nại nói trên được xác định như sau.

Đơn thư	Tuổi (x)	Gía trị SP (y)	Hạng x (R_{ix})	Hạng y (R_{iy})	R_{ix}-R_{iy} (d_i)	d_i^2
1	22	2,5	1	2	-1	1
2	26	1,8	6 (9)	1	8	64
3	25	12,5	2 (3,5)	18	-14,5	210,25
4	27	4,5	13 (14,5)	7	7,5	56,25
5	26	2,8	7 (9)	4	5	25
6	26	6,4	8 (9)	11	-2	4
7	25	10,2	3 (3,5)	17	-13,5	182,25
8	27	3,5	14 (14,5)	5	9,5	90,25
9	26	6,8	9	12	-3	9
10	48	5,5	17	9	8	64
11	26	4,7	10 (9)	8	1	1
12	25	8,2	4 (3,5)	14	-10,5	110,25
13	26	9,1	11 (9)	15	-6	36
14	25	7,4	5 (3,5)	13	-9,5	90,25
15	60	5,8	18	10	8	64
16	27	4,4	15 (14,5)	6	8,5	72,25
17	26	9,7	12 (9)	16	-7	49
18	27	2,6	16 (14,5)	3	11,5	132,25
Tổng	-	-	-	-	-	1261

(Số trong ngoặc là hạng trung bình của những tuổi bằng nhau)

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6 \times 1261}{18(18^2-1)} = -0{,}301$$

Vậy, giữa tuổi và giá trị sản phẩm mà khách hàng khiếu nại ở ví dụ trên có liên hệ tương quan nghịch rất yếu.

2.7. Các chỉ tiêu mô tả mối liên hệ giữa hai tiêu thức (biến) định tính

2.7.1. Mô tả mối liên hệ giữa hai tiêu thức danh định: Mối liên hệ giữa hai tiêu thức danh định có thể được mô tả tóm lược qua Hệ số Cramer và Hệ số liên hợp. Các hệ số này được xác định dựa vào bảng chéo (bảng phân tổ kết hợp) theo hai tiêu thức.

Trong đó:
* k : Số tổ của tiêu thức 1 (X)
* m : Số tổ của tiêu thức 2 (Y)
* f_{ij} : Tần số thực tế kết hợp của cả hai tiêu thức
* f_{ix} : Tần số riêng theo tiêu thức 1 (X)
* f_{jy} : Tần số riêng theo tiêu thức 2 (Y)

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Tiêu thức 1 (X)	Tiêu thức 2 (Y)				Cộng
	y_1	y_2	…	y_m	
x_1	f_{11}	f_{12}	…	f_{1m}	f_{1x}
x_2	f_{21}	f_{21}	…	f_{2m}	f_{2x}
…	…	…	…	…	…
x_k	f_{k1}	f_{k2}	…	f_{km}	f_{kx}
Cộng	f_{1y}	f_{2y}	…	f_{my}	n

Trong trường hợp hai tiêu thức độc lập, tần suất kết hợp của một ô bằng tích của hai tần suất riêng theo dòng và theo cột tương ứng của ô đó. Hay:

$$\frac{e_{ij}}{n} = \frac{f_{ix}}{n} \cdot \frac{f_{jy}}{n}$$

Do đó, tần số kết hợp kỳ vọng của ô ij khi hai tiêu thức x và y độc lập sẽ là:

$$e_{ij} = \frac{f_{ix} f_{jy}}{n}$$

- Hệ số Cramer V: $V = \sqrt{\dfrac{\chi^2}{n(h-1)}}$ $(0 \le V \le 1)$

Trong đó: h = min(k, m)

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

Ý nghĩa: + V = 0: Giữa hai tiêu thức không có mối liên hệ
+ V: Càng gần 0 giữa hai tiêu thức có mối liên hệ càng yếu
+ V: Càng gần 1 giữa hai tiêu thức có mối liên hệ càng mạnh

- Hệ số liên hợp (Coeficient of contingency):

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \qquad (0 \le C \le 1)$$

Ý nghĩa: * C = 0: Giữa hai tiêu thức không có mối liên hệ
* C: Càng gần 0 giữa hai tiêu thức có mối liên hệ càng yếu
* C: Càng gần 1 giữa hai tiêu thức có mối liên hệ càng mạnh

Ví dụ: Xét dữ liệu về 18 đơn thư khiếu nại ở trang 3. Mối liên hệ giữa hai tiêu thức Giới tính và Yêu cầu của khách hàng được mô tả như sau.

Phân tổ kết hợp và lập bảng chéo theo hai tiêu thức. Tính tần số kết hợp kỳ vọng khi 2 tiêu thức độc lập (số trong ngoặc của bảng) theo công thức:

$$e_{ij} = \frac{f_{ix} f_{jy}}{n}$$

Giới tính	Yêu cầu của khách hàng			Cộng
	Bồi thường	Đổi	Sửa	
Nam	1 (4)	3 (2,5)	5 (2,5)	9
Nữ	7 (4)	2 (2,5)	0 (2,5)	9
Cộng	8	5	5	18

(Số trong ngoặc là e_{ij})

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} = \frac{(1-4)^2}{4} + \frac{(3-2{,}5)^2}{2{,}5} + \ldots + \frac{(0-2{,}5)^2}{2{,}5} = 9{,}7$$

Hệ số Cramer: $V = \sqrt{\dfrac{\chi^2}{n(h-1)}} = \sqrt{\dfrac{9{,}7}{18(2-1)}} = 0{,}734$

Hệ số liên hợp: $C = \sqrt{\dfrac{\chi^2}{\chi^2 + n}} = \sqrt{\dfrac{9{,}7}{9{,}7+18}} = 0{,}592$

Cả hai hệ số là khá lớn cho thấy mối liên hệ khá rõ giữa giới tính và loại yêu cầu của các khách hàng khi khiếu nại. Cụ thể, nữ giới có phần thiên về đòi bồi thường trong khi nam giới có phần thiên về yêu cầu sửa hoặc đổi.

2.7.2. Mô tả mối liên hệ của hai tiêu thức thứ bậc: Mối liên hệ giữa hai tiêu thức thứ bậc có thể được mô tả tóm lược qua Hệ số Gamma, Hệ số Kendall-Tau.

- Hệ số Gammar: $\gamma = \dfrac{n_c - n_d}{n_c + n_d}$ $\gamma \in (-1, 1)$

- Hệ số Kendall-Tau A: $\tau_A = 2\dfrac{n_c - n_d}{n(n-1)}$ $\tau_A \in (-1, 1)$

Trong đó: n: Tổng số quan sát (số đơn vị tổng thể của dữ liệu)
n_c: Số cặp quan sát có sự đồng biến trên hai tiêu thức
n_d: Số cặp quan sát có sự nghịch biến trên hai tiêu thức

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Để xác định số cặp đồng biến và nghịch biến được thuận lợi, người ta thường dựa vào bảng chéo (phân tổ kết hợp) theo hai tiêu thức.

Ý nghĩa: * $\gamma = 0,\ \tau_A = 0$: Giữa hai tiêu thức không có liên hệ

* $\gamma > 0,\ \tau_A > 0$: Giữa hai tiêu thức có liên hệ thuận

* $\gamma < 0,\ \tau_A < 0$: Giữa hai tiêu thức có liên hệ nghịch

* γ, τ_A : Càng gần 0 giữa hai tiêu thức có liên hệ càng yếu

* γ, τ_A : Càng gần ±1 giữa hai tiêu thức có liên hệ càng mạnh

Ngoài ra người ta còn có thể dùng Hệ số Kendall-Tau B, Kendall-Tau C có cách tính có phần tương tự nhưng phức tạp hơn. Những hệ số này thường được sử dụng trong các phần mềm thống kê như SPSS.

Ví dụ: Để đánh giá mối liên hệ giữa mức độ biết đến thương hiệu và mức độ ưa thích sản phẩm của một doanh nghiệp, 200 người tiêu dùng được chọn ngẫu nhiên để thăm dò ý kiến. Kết quả được phân tổ kết hợp trong bảng chéo như sau.

Mức độ ưa thích sản phẩm	Mức độ biết đến thương hiệu		
	Kém	Khá	Tốt
Thấp	50	10	10
Trung bình	20	40	20
Cao	0	10	40

Số cặp quan sát có sự đồng biến trên hai tiêu thức được xác định như sau:

50 →		
	40	20
	10	40

$n_{c(50)}$ = 50(40+20+10+40) = 5500

10 →	
	20
	40

$n_{c(10)}$ = 10(20+40) = 600

20 →	
10	40

$n_{c(20)}$ = 20(10+40) = 1000

40 →	
	40

$n_{c(40)}$ = 40*40 = 1600

=> n_c = 5500 + 600 + 1000 + 1600 = 8700

Số cặp quan sát có sự nghịch biến trên hai tiêu thức được xác định như sau:

	←	10
20	40	
0	10	

$n_{d(10)}$ = 10(40+20+10+0) = 700

←	10
20	
0	

$n_{d(10)}$ = 10(20+0) = 200

	←	20
0	10	

$n_{d(20)}$ = 20(10+0) = 200

←	40
0	

$n_{d(40)}$ = 40*0 = 0

=> n_d = 700 + 200 + 200 + 0 = 1100

- Hệ số Gammar: $\gamma = \frac{n_c - n_d}{n_c + n_d} = \frac{8700 - 1100}{8700 + 1100} = 0,77$

- Hệ số Kendall-Tau A: $\tau_A = 2\frac{n_c - n_d}{n(n-1)} = 2\frac{8700 - 1100}{200(200-1)} = 0,38$

Các hệ số này cho thấy có mối liên hệ thuận chiều đáng kể giữa mức độ biết đến thương hiệu và mức độ ưa thích sản phẩm của doanh nghiệp.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Chương 3 **PHÂN PHỐI XÁC SUẤT CỦA BIẾN NGẪU NHIÊN**

3.1. Biến ngẫu nhiên

3.1.1. Khái niệm biến ngẫu nhiên

Một biến ngẫu nhiên là một cách thức mô tả kết quả của một phép thử ngẫu nhiên dưới dạng các con số.

Ví dụ: Tuổi của một sinh viên được chọn ngẫu nhiên trong một lớp. Sai số chiều dài một sản phẩm được chọn ngẫu nhiên so với tiêu chuẩn.

Biến ngẫu nhiên chỉ nhận một số đếm được các giá trị gọi là biến ngẫu nhiên rời rạc. Biến ngẫu nhiên có khả năng nhận mọi giá trị trên một khoảng của trục số gọi là biến ngẫu nhiên liên tục.

Biến ngẫu nhiên thường được ký hiệu là X, Y, Z … Các giá trị của nó thường được ký hiệu tương ứng là $x_1, x_2, x_3 \ldots, y_1, y_2, y_3 \ldots, z_1, z_2, z_3 \ldots$

3.1.2. Phân phối xác suất của biến ngẫu nhiên (Probability distributions)

Một hình thức xác lập mối quan hệ giữa các giá trị và các xác suất tương ứng của một biến ngẫu nhiên gọi là phân phối xác suất của biến ngẫu nhiên ấy.

Người ta thường dùng một trong các phương pháp bảng, biểu đồ hay biểu thức đại số để xác lập phân phối xác suất của biến ngẫu nhiên.

Ví dụ, có bảng thống kê tuổi của sinh viên trong một lớp như sau:

Tuổi	19	20	21	22	23	24
Tần suất (%)	10	20	50	10	6	4

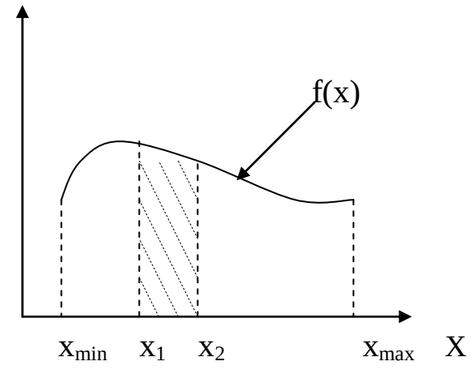
Phân phối tuổi (X) của một sinh viên được chọn ngẫu nhiên trong lớp trên có thể được trình bày trong bảng sau:

x_i	19	20	21	22	23	24
P_i	0,1	0,2	0,5	0,1	0,06	0,04

Đối với biến ngẫu nhiên liên tục, ta không dùng được phương pháp bảng như trên. Thay vào đó, người ta thường dùng hàm mật độ phân phối để trình bày luật phân phối. Hàm mật độ phân phối f(x) của biến ngẫu nhiên X có những tính chất sau:

* $f(x) \geq 0 \qquad \forall x$

* Trên đồ thị, xác suất mà $x_1 \leq X \leq x_2$ là diện tích hình nằm bên dưới hàm mật độ phân phối, giữa x_1 và x_2.



Hay: $$p(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)dx$$

* $$p(x_{min} \leq X \leq x_{max}) = \int_{x_{min}}^{x_{max}} f(x)dx = 1$$

3.2. Các tham số đặc trưng của biến ngẫu nhiên

3.2.1. Kỳ vọng toán E(X) (Expected value):

- Kỳ vọng toán biến ngẫu nhiên rời rạc:

$$E(X) = \sum_{i=1}^{n} x_i p_i$$

Ví dụ: Xét biến ngẫu nhiên (X) về tuổi của sinh viên được chọn ở trên.

x_i	19	20	21	22	23	24
P_i	0,1	0,2	0,5	0,1	0,06	0,04

Kỳ vọng toán của X được xác định như sau:

$$E(X) = \sum_{i=1}^{n} x_i p_i = 19.0{,}1 + 20.0{,}2 + \ldots + 24.0{,}04 = 20{,}9$$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Kỳ vọng toán chính là số đo trung tâm của biến ngẫu nhiên nên cũng được gọi là Số trung bình và thường được ký hiệu gọn hơn là μ.

- Kỳ vọng toán biến ngẫu nhiên liên tục:

$$E(X) = \int_{x_{min}}^{x_{max}} x\, f(x)\, dx$$

3.2.2. Phương sai V(X) (Variance):

- Phương sai biến ngẫu nhiên rời rạc:

$$V(X) = \sum_{i=1}^{n} [x_i - E(X)]^2 p_i$$

Ví dụ: Xét biến ngẫu nhiên (X) về tuổi của sinh viên được chọn ở trên.

x_i	19	20	21	22	23	24
P_i	0,1	0,2	0,5	0,1	0,06	0,04

Phương sai của X được xác định như sau:

$$V(X) = \sum_{i=1}^{n} [x_i - E(X)]^2 p_i$$

$$= (19\text{-}20{,}9)^2.0{,}1+(20\text{-}20{,}9)^2.0{,}2+ \ldots +(24\text{-}20{,}9)^2.0{,}04 = 1{,}3$$

Trong đó: $E(X) = \sum_{i=1}^{n} x_i p_i = 19.0{,}1 + 20.0{,}2 + \ldots + 24.0{,}04 = 20{,}9$

Phương sai chính là số đo đo lường độ phân tán của biến ngẫu nhiên nên cũng thường được ký hiệu là σ^2.

- Phương sai biến ngẫu nhiên liên tục:

$$V(X) = \int_{x_{min}}^{x_{max}} [x - E(X)]^2\, f(x)\, dx$$

3.2.3. Hiệp phương sai (Covariance):

Hiệp phương sai của hai biến ngẫu nhiên x và y được xác định theo công thức:

$$Cov(x, y) = E[(x - E(x))(y - E(y))]$$

Trường hợp hai biến ngẫu nhiên rời rạc:

$$Cov(x, y) = \sum_{i,j} (x_i - E(x))(y_j - E(y)) p_{ij}$$

Trong đó p_{ij} là xác suất xuất hiện đồng thời cặp giá trị x_i và y_j.

Ví dụ: Có phân phối xác suất kết hợp của hai biến ngẫu nhiên: Số sản phẩm mua trong năm (X) và Số lần khiếu nại (Y) như sau.

Số sản phẩm mua trong năm (x_i)	Số lần khiếu nại (y_j)			P_i
	0	1	2	
1	0,50	0,16	0,08	0,74
2	0,20	0,04	0,02	0,26
P_j	0,70	0,20	0,10	1

$$E(X) = \sum_{i=1}^{n} x_i p_i = 1.0{,}74 + 2.0{,}26 = 1{,}26$$

$$E(Y) = \sum_{j=1}^{n} y_j p_j = 0.0{,}70 + 1.0{,}20 + 2.0{,}10 = 0{,}4$$

Việc tính toán hiệp phương sai có thể thực hiện dựa vào bảng sau.

x_i	y_j	p_{ij}	x_i-E(x)	y_j-E(y)	$(x_i\text{-}E(x))(y_j\text{-}E(y))p_{ij}$
1	0	0,50	-0,26	-0,4	0,052
1	1	0,16	-0,26	0,6	-0,025
1	2	0,08	-0,26	1,6	-0,033
2	0	0,20	0,74	-0,4	-0,059
2	1	0,04	0,74	0,6	0,018
2	2	0,02	0,74	1,6	0,024
Tổng					-0,024

$$Cov(x, y) = \sum_{i,j} (x_i - E(x))(y_j - E(y)) p_{ij} = -0{,}024$$

Hiệp phương sai đo lường mối liên hệ tương quan tuyến tính giữa hai biến ngẫu nhiên:

+ Cov(x,y) = 0 giữa hai biến ngẫu nhiên x, y không có liên hệ tương quan.

+ Cov(x,y) > 0 giữa hai biến ngẫu nhiên x, y có liên hệ tương quan thuận chiều.

+ Cov(x,y) < 0 giữa hai biến ngẫu nhiên x, y có liên hệ tương quan nghịch chiều.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

3.3. Một số phân phối cơ bản thường dùng trong thống kê

3.3.1. Phân phối Bernoulli X ~ A(p)

- *Định nghĩa:* Biến ngẫu nhiên rời rạc X có phân phối Bernoulli khi nó chỉ có hai giá trị 0 và 1 với xác suất của giá trị 1 là p và giá trị 0 là 1-p.

- *Bảng phân phối:* Với p là hằng số, 0 < p < 1

X	0	1
P(X)	1-p	p

- *Đặc điểm:* * E(X) = p
* V(X) = p(1-p)

- *Vận dụng thực tế:* Đây là phân phối của một phép thử ngẫu nhiên chỉ có hai kết quả và xác suất xuất hiện của một kết quả nào đó (X=1) là p còn xác suất xuất hiện của kết quả còn lại (X=0) sẽ là 1-p.

Ví dụ: Một kỳ kế toán của một doanh nghiệp có 200 bút toán. Trong đó có 10 bút toán có vấn đề. Chọn ngẫu nhiên một bút toán để kiểm tra.

Gọi X là biến chỉ vấn đề của bút toán được chọn, với X = 1 chỉ bút toán có vấn đề, X = 0 chỉ bút toán không có vấn đề.

P(X = 1) = p = 10/200 = 0,05 (kỳ vọng bút toán có vấn đề)

P(X = 0) = 1- p = 0,95

Do đó, X có phân phối Bernoulli X~ A(0,05):

X	0	1
P(X)	0,95	0,05

E(X) = p = 0,05

Var(X) = p(1-p) = 0,05(1-0,05)= 0,0475

3.3.2. Phân phối nhị thức (Binomial distribution) X~ B(n, p)

- *Định nghĩa:* Biến ngẫu nhiên rời rạc X có phân phối nhị thức khi nó có (n+1) giá trị: 0, 1, 2, …, n với xác suất nhận giá trị x là:

$$P_{(x)} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$ (p là hằng số, 0 < p < 1)

Bảng phân phối nhị thức thường được lập sẵn theo các thông số n, p và x khác nhau để phục vụ việc tra cứu (xem phụ lục).

- *Đặc điểm:* * E(X) = np , V(X) = np(1-p)

* Phân phối nhị thức xấp xỉ phân phối chuẩn khi np≥5 và n(1-p)≥5 .

- *Vận dụng thực tế:* Đây là phân phối của số lần xuất hiện một kết quả quan tâm nào đó (gọi là thành công) trong dãy n phép thử ngẫu nhiên với xác suất của kết quả quan tâm (thành công) là p, không đổi trong mọi phép thử.

Ví dụ: Một ứng viên việc làm thực hiện một bài thi trắc nghiệm gồm 8 câu. Mỗi câu có 5 đáp án, trong đó chỉ có một đáp án đúng. Giả sử rằng ứng viên này hoàn toàn không biết đáp án nào đúng, chỉ chọn ngẫu nhiên các đáp án.

Việc trả lời 8 câu hỏi của ứng viên trên có thể được xem như 8 phép thử ngẫu nhiên có cùng xác suất trả lời đúng (thành công) ở mỗi phép thử (mỗi câu) là p = 1/5 = 0,2 (không đổi trong cả 8 câu).

Gọi X là biến chỉ số câu trả lời đúng (thành công) của ứng viên nói trên. X sẽ có phân phối nhị thức X~ B(8, 0,2):

X	0	1	2	3	4	5	6	7	8
P(X)	0,168	0,336	0,294	0,147	0,046	0,009	0,001	0,000	0,000

(Xem bảng phân phối nhị thức ở phụ lục)

E(X) = n.p = 8.0,2 = 1,6

Var(X) = np(1-p) = 8.0,2.(1-0,2)= 1,28

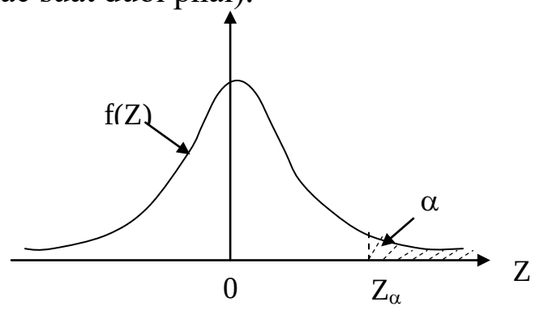
3.3.3. Phân phối chuẩn tắc (Standard normal distribution) Z ~ N(0, 1)

- *Định nghĩa:* Biến ngẫu nhiên liên tục Z ∈ (-∞,+∞) có phân phối chuẩn tắc khi hàm mật độ phân phối là:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

- *Đặc điểm:* * E (Z) = 0
* V (Z) = 1
* Đồ thị phân phối hình quả chuông, đối xứng qua trục tung.

- *Bảng phân vị chuẩn:* Để tiện cho việc tra cứu, người ta lập bảng phân vị chuẩn thể hiện mối quan hệ giữa giá trị Z_α với xác suất α mà Z lấy tất cả các giá trị từ Z_α đến +∞ (xác suất đuôi phải).



Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Z_α được gọi là điểm phân vị (cutoff point) mức α, còn α gọi là giá trị xác suất đuôi phải (P-value) của điểm phân vị Z_α. Ký hiệu: $\alpha = \text{P-value}_{(Z\alpha)}$.

Một số trị số thường dùng của bảng phân vị chuẩn:

α	0,005	0,01	0,02	0,025	0,05	0,1
Z_α	2,575	2,326	2,055	1,960	1,645	1,28

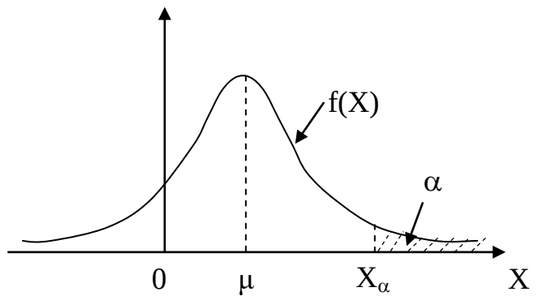
α	0,995	0,99	0,977	0,975	0,95	0,9
Z_α	-2,575	-2,326	-2,000	-1,960	-1,645	-1,28

Bảng phân vị chuẩn có tính đối xứng: $Z_{1-\alpha} = -Z_\alpha$.

3.3.4. Phân phối chuẩn (Normal distribution) X~ N(μ,σ²)

- *Định nghĩa:* Biến ngẫu nhiên liên tục X ∈ (-∞,+∞) có phân phối chuẩn khi hàm mật độ phân phối là:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



- *Đặc điểm:* * E (X) = μ
  * V (X) = σ^2

* Đồ thị phân phối hình quả chuông, đối xứng qua đường thẳng X = μ.

* Phân phối chuẩn dễ dàng được chuyển về phân phối chuẩn tắc bằng cách đặt: $Z = \frac{X-\mu}{\sigma}$. Do đó, ta có thể sử dụng bảng phân vị chuẩn nói trên để tra cứu các phân vị của phân phối chuẩn nói chung.

Ví dụ, xét biến ngẫu nhiên X~ N(28, 2^2) => $Z = \frac{X-\mu}{\sigma} = \frac{X-28}{2} \sim N(0,1)$

* Nếu X_α = 30: => $Z_\alpha = \frac{X_\alpha - \mu}{\sigma} = \frac{30-28}{2} = 1$

Tra bảng phân vị chuẩn, ta được xác suất đuôi phải: α = 0,16.

* Nếu xác suất đuôi phải α = 0,025, ta có thể tra điểm phân vị X_α của nó:

Tra bảng phân vị chuẩn, ta được $Z_\alpha = 1,96$

$$=> X_\alpha = Z_\alpha\sigma + \mu = 1,96.2 + 5 = 8,92$$

- Định lý giới hạn trung tâm (The central limit theorem):

Nếu $X_1, X_2, \ldots, X_n$ là n biến ngẫu nhiên độc lập, có cùng phân phối bất kỳ, cùng kỳ vọng toán μ và cùng phương sai σ^2 thì biến ngẫu nhiên trung bình: $\overline{X} = \frac{\sum X_i}{n}$ có phân phối chuẩn, khi n đủ lớn.

Cụ thể: $$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Với điều kiện: n ≥ 30 nếu các X_i có cùng phân phối bất kỳ.
n ≥ 15 nếu các X_i có cùng phân phối đối xứng.
n bất kỳ nếu các X_i có cùng phân phối chuẩn.

- *Ý nghĩa vận dụng:* Chọn lặp ngẫu nhiên từ một tổng thể sẽ cho các biến thỏa mãn các điều kiện của định lý giới hạn trung tâm nên trung bình của chúng có phân phối chuẩn khi kích thước mẫu đủ lớn.

Ví dụ: Một doanh nghiệp có 10.000 CN với tuổi nghề trung bình là 28, phương sai tuổi nghề là 128. Chọn ngẫu nhiên, lặp 32 công nhân từ doanh nghiệp này.

Gọi X_i là tuổi nghề công nhân thứ i được chọn. Các X_i độc lập nhau, có cùng phân phối với phân phối tuổi nghề của doanh nghiệp, có cùng tuổi nghề trung bình là 28 và có cùng phương sai tuổi nghề là 128.

Các X_i thỏa mãn điều kiện của định lý giới hạn trung tâm. Do đó tuổi nghề trung bình của 32 công nhân được chọn sẽ có phân phối chuẩn:

$$\overline{X} \sim N(28, \frac{128}{32}) \quad \text{hay} \quad \overline{X} \sim N(28, 2^2)$$

Sử dụng bảng phân vị chuẩn ta có thể tra cứu xác xuất mà tuổi nghề trung bình của 32 CN có giá trị từ 30 tuổi trở lên (xác suất đuôi phải) là α = 0,16 (thực hiện tương tự ví dụ kế trên)

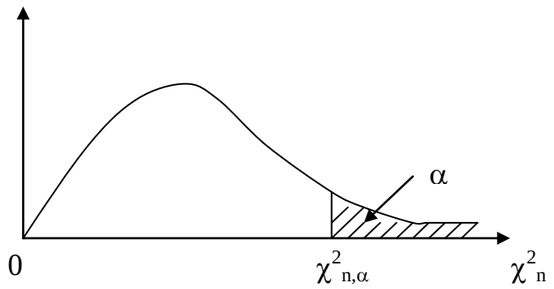
3.3.5. Phân phối Khi bình phương (Chi square distribution) X ~ χ^2_n

- *Định nghĩa:* Biến ngẫu nhiên liên tục X ∈ (0,+∞) có phân phối Khi bình phương khi có hàm mật độ phân phối là:

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

$$f(x)=\frac{x^{(\frac{n}{2}-1)}e^{\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}$$



Trong đó: * $\Gamma(u)=\int_0^{+\infty} v^{u-v}e^{-v}dv$: Hàm Gamma

* n: Bậc tự do (hằng số nguyên, dương)

- *Đặc điểm:* $E(\chi^2_n) = n$

- *Bảng phân vị Khi bình phương:* Để tiện cho việc tra cứu, người ta lập bảng phân vị thể hiện mối quan hệ giữa giá trị $\chi^2_{n,\alpha}$ với xác suất α mà χ^2_n lấy tất cả các giá trị từ $\chi^2_{n,\alpha}$ đến $+\infty$ (xác suất đuôi phải).

- *Ý nghĩa vận dụng:*

* Nếu X_i với $i=\overline{1,n}$ là n biến ngẫu nhiên độc lập, có cùng phân phối chuẩn, cùng kỳ vọng toán và cùng phương sai (σ^2), thì:

$$\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}\sim\chi^2_{n-1}$$

* Giả sử có n biến ngẫu nhiên độc lập, có cùng phân phối chuẩn, cùng kỳ vọng toán và cùng phương sai (σ^2). Nếu n biến này được chia thành t nhóm, nhóm i có n_i biến, thì:

$$\frac{\sum_{i=1}^{t}(\overline{X}_i-\overline{X})^2 n_i}{\sigma^2}\sim\chi^2_{t-1}$$

Trong đó:

$\overline{X}_i$: Số trung bình của các biến ngẫu nhiên trong nhóm i

$\overline{X}$: Số trung bình của tất cả n biến ngẫu nhiên

* Tổng của các biến Khi bình phương là một biến Khi bình phương có bậc tự do bằng tổng bậc tự do của các biến đó.

$$\sum_{i=1}^{t}\chi^2_{n_i}\sim\chi^2_n \quad \text{với} \quad n=\sum_{i=1}^{t}n_i$$

3.3.6. Phân phối Student $T \sim T_n$

- *Định nghĩa:* Biến ngẫu nhiên liên tục $T \in (-\infty,+\infty)$ có phân phối Student khi hàm mật độ phân phối là:

$$f(t)=\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\sqrt{(n-1)\pi}}(1+\frac{t^2}{n-1})^{-\frac{n}{2}}$$

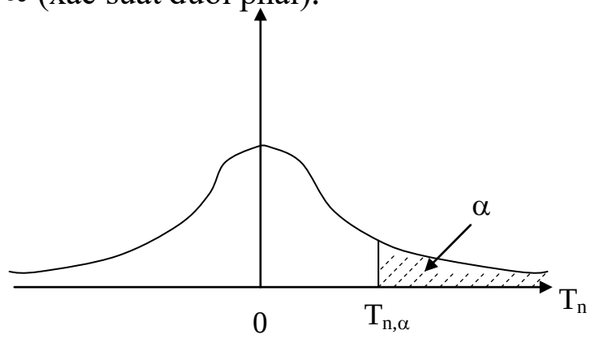
Trong đó n là bậc tự do (hằng số nguyên, dương)

- *Đặc điểm:* * E(T) = 0.

* Đồ thị phân phối hình quả chuông, đối xứng qua trục tung, tương tự phân phối chuẩn tắc nhưng phân tán hơn.

* Khi n ≥30 phân phối Student xấp xỉ phân phối chuẩn tắc.

- *Bảng phân vị Student:* Để tiện cho việc tra cứu, người ta lập bảng phân vị Student thể hiện mối quan hệ giữa giá trị $T_{n,\alpha}$ với xác suất α mà T_n lấy tất cả các giá trị từ $T_{n,\alpha}$ đến $+\infty$ (xác suất đuôi phải).



Bảng phân vị này cũng có tính đối xứng: $T_{n,1-\alpha}=-T_{n,\alpha}$.

- *Ý nghĩa vận dụng:* Nếu X_i với $i=\overline{1,n}$ là n biến ngẫu nhiên thỏa mãn định lý giới hạn trung tâm để $\overline{X}$ có phân phối chuẩn thì:

$$\frac{\overline{X}-\mu}{\hat{\sigma}_{\overline{X}}}\sim T_{n-1}$$

Trong đó $\hat{\sigma}_{\overline{X}}$ là một ước lượng không chệch của $\sigma_{\overline{X}}$ hay $E(\hat{\sigma}_{\overline{X}})=\sigma_{\overline{X}}$.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

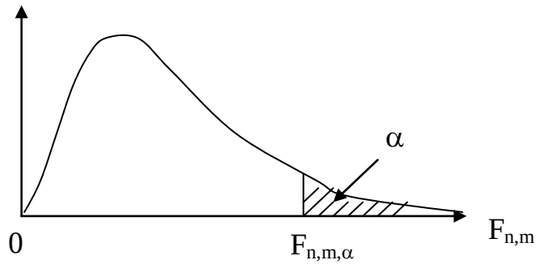
3.3.7. Phân phối Fisher-Snedecor $X \sim F_{n,m}$

- *Định nghĩa:* Biến ngẫu nhiên liên tục $X \in (0,+\infty)$ có phân phối Fisher-Snedecor khi có hàm mật độ phân phối là:

$$f(x) = \frac{\Gamma(\frac{n+m}{2}) n^{\frac{n}{2}} m^{\frac{m}{2}}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \cdot \frac{x^{\frac{n-m}{2}}}{(m+nx)^{\frac{n+m}{2}}}$$

Trong đó n, m là các bậc tự do (hằng số nguyên, dương)

- *Bảng phân vị Fisher:* Để tiện cho việc tra cứu, người ta lập bảng phân vị thể hiện mối quan hệ giữa giá trị $F_{n,m,\alpha}$ với xác suất α mà $F_{n,m}$ lấy tất cả các giá trị từ $F_{n,m,\alpha}$ đến $+\infty$ (xác suất đuôi phải).



- *Ý nghĩa vận dụng:* Nếu χ^2_{n1} và χ^2_{n2} là hai biến ngẫu nhiên độc lập, có phân phối Khi bình phương với n_1 và n_2 bậc tự do, thì:

$$F = \frac{\frac{\chi^2_{n_1}}{n_1}}{\frac{\chi^2_{n_2}}{n_2}} \sim F_{n_1,n_2}$$

Chương 4 LẤY MẪU VÀ ƯỚC LƯỢNG THAM SỐ TỔNG THỂ

Trong thực tế cuộc sống, chúng ta thường xuyên làm việc bằng cách lấy mẫu để thử rồi suy rộng. Chẳng hạn, để biết một quán ăn nào đó thức ăn (tổng thể) có ngon không, chúng ta thường đến ăn thử một bữa (mẫu). Để quyết định chọn mua một loại gạo (tổng thể) chúng ta bốc một nắm (mẫu) để xem. Để biết nước tắm (tổng thể) có vừa đủ ấm, chúng ta thường cho đầu ngón tay tiếp xúc với một ít nước (mẫu)… Tuy nhiên, *trong quản lý kinh tế - xã hội, lấy mẫu thế nào để bảo đảm độ tin cậy của sự suy rộng từ mẫu cho tổng thể là một vấn đề đặt ra và cần phải được giải quyết một cách khoa học.*

4.1. Lấy mẫu

4.1.1. Mẫu ngẫu nhiên đơn giản và mẫu ngẫu nhiên

Khi lấy mẫu từ một tổng thể hữu hạn, các nhà thống kê thường chọn cách lấy mẫu xác suất bởi vì một mẫu xác suất cho phép thực hiện các suy luận thống kê đáng tin cậy về các tham số tổng thể từ các thống kê mẫu.

Loại mẫu xác suất đơn giản nhất là loại mẫu, theo đó mỗi mẫu kích thước n có xác suất được chọn ra như nhau. Ví dụ: Xét tổng thể có 5 phần tử là A, B, C, D, E.

Nếu được phép lấy lặp, có tất cả 15 mẫu kích thước n=2 bao gồm: AA, BB, CC, DD, EE, AB, AC, AD, AE, BC, BD, BE, CD, CE, DE. Một trong 15 mẫu này là mẫu ngẫu nhiên đơn giản với xác suất được chọn như nhau là 1/15.

Nếu lấy không lặp, có tất cả 10 mẫu kích thước n=2 bao gồm: AB, AC, AD, AE, BC, BD, BE, CD, CE, DE. Một trong 10 mẫu này là mẫu ngẫu nhiên đơn giản với xác suất được chọn như nhau là 1/10.

Mẫu ngẫu nhiên đơn giản *kích thước n từ một tổng thể hữu hạn kích thước N là mẫu được chọn sao cho mọi mẫu có cùng kích thước n đều có xác suất được chọn như nhau.*

Đôi khi tổng thể lấy mẫu là tổng thể vô hạn. Đó là tổng thể lớn vô hạn hay các đơn vị tổng thể đang được sinh ra thêm một cách vô hạn. Ví dụ: Các bộ phận đang được sản xuất trên một dây chuyền sản xuất, các giao dịch đang xảy ra tại một ngân hàng, các cuộc gọi đến một tổ hỗ trợ kỹ thuật, các khách hàng đang đi vào một cửa hàng. Trong trường hợp này không thể lấy mẫu ngẫu nhiên đơn giản được, các nhà thống kê thường dùng mẫu ngẫu nhiên.

Mẫu ngẫu nhiên *kích thước n từ một tổng thể vô hạn là mẫu được chọn sao cho các điều kiện sau được thỏa mãn:*

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

(1) *Mỗi đơn vị tổng thể của mẫu được chọn từ một tổng thể như nhau.*

(2) *Mỗi đơn vị tổng thể của mẫu được chọn một cách độc lập.*

4.1.2. Lấy mẫu lặp và lấy mẫu không lặp

Lấy mẫu lặp là thủ tục lấy mẫu cho phép hai hay nhiều đơn vị tổng thể trong cùng một mẫu có thể trùng nhau.

Lấy mẫu không lặp là thủ tục lấy mẫu không cho phép hai hay nhiều đơn vị tổng thể trong cùng một mẫu có thể trùng nhau.

Cả hai cách lấy mẫu trên đều cho các mẫu ngẫu nhiên đơn giản từ tổng thể hữu hạn. Tuy nhiên, *lấy mẫu không lặp là thủ tục lấy mẫu thường được dùng nhất trong thực hành.* Vì vậy, khi nói đến lấy mẫu ngẫu nhiên đơn giản là nói đến lấy mẫu không lặp.

4.1.3. Phương pháp thực hành lấy mẫu ngẫu nhiên

Thực hành lấy mẫu ngẫu nhiên đơn giản từ tổng thể hữu hạn:

- Bước 1, lập khung lấy mẫu (dàn mẫu): Để lấy một mẫu ngẫu nhiên đơn giản từ một tổng thể hữu hạn, cần xây dựng khung lấy mẫu. Khung lấy mẫu là một danh sách liệt kê một đặc điểm nhận dạng nào đó của tất cả các đơn vị tổng thể được dùng làm căn cứ để lấy mẫu.

Ví dụ, khung lấy mẫu của tổng thể gồm 200 khách hàng của một doanh nghiệp là một danh sách tên 200 khách hàng hoặc danh sách địa chỉ của 200 khách hàng này.

- Bước 2, đánh số các phần tử trên khung: Ví dụ, ta có khung lấy mẫu gồm tên của 200 khách hàng. Vì N=200 có ba chữ số, ta đánh số cho mỗi tên khách hàng trong khung lấy mẫu từ 001, 002,…, cho đến 200.

- Bước 3, lấy các phần tử từ khung: Để lấy ra các phần tử từ khung lấy mẫu, người ta thường dùng cách quay số ngẫu nhiên. Hàm ngẫu nhiên RANDBETWEEN của Excel có thể được sử dụng. *Ví dụ*, ta cần chọn mẫu ngẫu nhiên đơn giản 20 khách hàng từ khung lấy mẫu gồm tên của 200 khách hàng. Hàm ngẫu nhiên RANDBETWEEN của Excel với chỉ định Bottom = 1 và Top = 200 cho ta 20 số ngẫu nhiên như sau.

158	94	19	3	31	28	104	45	85	197
123	173	145	28	170	177	138	139	178	168

Các phần tử trên khung lấy mẫu mang các số trong bảng trên được lấy làm mẫu ngẫu nhiên đơn giản.

Chú ý: Có thể sử dụng bảng số ngẫu nhiên (xem phụ lục) để lấy các phần tử từ khung như sau. Ta dùng ba chữ số cuối của mỗi cột số để chọn, nhóm ba chữ số cuối nào vượt quá 200 thì bỏ qua. Lần lượt chọn cho đến khi đủ 20 số tương đương với 20 khách hàng thì dừng lại. Chẳng hạn trong cột đầu ta chọn được các số: 007, 022, 073; cột thứ hai ta chọn được các số: 011, 059, 110, 100, 076, 129, 109, 030; tiếp tục ở cột thứ ba, thứ tư… cho đến khi đủ 20 số.

Trên thực tế chỉ lấy mẫu không lặp nên những số chọn sau trùng với số đã được chọn trước đó thì bỏ qua để chọn số tiếp theo.

Thực hành lấy mẫu ngẫu nhiên từ tổng thể vô hạn:

Trong trường hợp này nói chung ta không thể xây dựng được khung lấy mẫu. Việc lấy mẫu đòi hỏi sự cẩn trọng và sự phán xét để bảo đảm thỏa mản hai điều kiện của một mẫu ngẫu nhiên từ một tổng thể vô hạn. Mỗi tình huống cụ thể đòi hỏi một thủ tục lấy mẫu phù hợp. Nói chung để thỏa màn điều kiện đầu, cần giới hạn tổng thể lấy mẫu về thời gian (hoặc không gian) nhằm tránh sự biến đổi đáng kể về tổng thể. Để thỏa màn điều kiện thứ hai, cần lấy các phần tử rải đều về thời gian (hoặc không gian), tránh cùng nhóm, cùng loại, ...

Ví dụ 1, lấy mẫu ngẫu nhiên 20 sản phẩm từ một dây chuyền sản xuất đang hoạt động. Ta không thể xây dựng khung lấy mẫu bao gồm tất cả các sản phẩm của dây chuyền, bởi vì dây chuyền hoạt động liên tục. Kết thúc công đoạn này sản phẩm đi ngay vào công đoạn khác.

Để thỏa mản điều kiện thứ nhất của mẫu ngẫu nhiên (tổng thể như nhau), các sản phẩm của mẫu cần được chọn trong một thời gian không quá dài để các điều kiện sản xuất của dây chuyền không thay đổi, một ca sản xuất chẳng hạn.

Để thỏa mản điều kiện thứ hai của mẫu ngẫu nhiên (mỗi đơn vị tổng thể được chọn độc lập), dây chuyền sản xuất phải được thiết kế sao cho việc sản suất các sản phẩm trên dây chuyền là độc lập nhau, chọn các sản phẩm với khoảng cách thời gian bằng nhau chẳng hạn.

Ví dụ 2, lấy mẫu ngẫu nhiên 30 khách hàng từ các khách hàng đến một nhà hàng ăn nhanh trong một ngày để phỏng vấn. Ta không thể xây dựng khung lấy mẫu bao gồm tất cả các khách hàng đến nhà hàng từ trước khi họ đến để lấy mẫu và biết được các vị khách nào cần phỏng vấn khi họ đến.

Để thỏa mản điều kiện thứ nhất của mẫu ngẫu nhiên (tổng thể như nhau), các khách hàng của mẫu cần được chọn từ những khách hàng đi vào nhà hàng và có mua thức ăn nhanh.

Để thỏa mản điều kiện thứ hai của mẫu ngẫu nhiên (mỗi đơn vị tổng thể được chọn độc lập), có thể chọn khách hàng kế tiếp một khách hàng có phiếu giảm giá vừa đi vào nhà hàng, hay chọn các khách hàng với khoảng cách thời gian cách đều nhau. Không nên chọn khách hàng theo một đặc điểm nào đó (chẳng hạn cùng độ tuổi hay cùng giới tính…) hay chọn theo từng nhóm khách hàng đi cùng nhau.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

4.2. Tham số tổng thể và thống kê mẫu

4.2.1. Các tham số tổng thể cơ bản

- Số trung bình tổng thể: $\mu = \frac{\sum x_i}{N}$ (x_i *tiêu thức định lượng)*

- Phương sai tổng thể: $\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$ (x_i *tiêu thức định lượng)*

- Tỉ lệ tổng thể: $p = \frac{X}{N}$

Trong đó X là số đơn vị tổng thể có biểu hiện nghiên cứu trên tổng thể.

4.2.2. Các thống kê mẫu cơ bản

+ Số trung bình mẫu: $\overline{x} = \frac{\sum x_i}{n}$

+ Phương sai mẫu: $s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$

+ Tỉ lệ mẫu: $\overline{p} = \frac{x}{n}$

Trong đó x là số đơn vị tổng thể có biểu hiện nghiên cứu trên mẫu.

4.3. Phân phối lấy mẫu

Phân phối lấy mẫu của một thống kê mẫu là phân phối xác suất của tất cả các giá trị có khả năng của thống kê mẫu đó trên vô số mẫu ngẫu nhiên cùng kích thước được lấy từ một tổng thể.

4.3.1. Phân phối lấy mẫu của Số trung bình mẫu:

Phân phối lấy mẫu của Số trung bình mẫu là phân phối xác suất của tất cả các giá trị có khả năng của số trung bình mẫu.

Lấy mẫu ngẫu nhiên đơn giản từ tổng thể vô hạn hoàn toàn đáp ứng các điều kiện của định lý giới hạn trung tâm đề cập ở Chương 3. Do đó, định lý giới hạn trung tâm có thể phát biểu cho phân phối lấy mẫu như sau: "Khi lấy mẫu ngẫu nhiên đơn giản kích thước n từ một tổng thể vô hạn, phân phối lấy mẫu của trung bình mẫu xấp xỉ phân phối chuẩn nếu kích thước mẫu đủ lớn".

- Trường hợp tổng thể vô hạn:

$$\overline{x} \sim N(\mu, \frac{\sigma^2}{n})$$

Với điều kiện:

+ $n \geq 30$ nếu tổng thể có phân phối bất kỳ
+ $n \geq 15$ nếu tổng thể có phân phối đối xứng
+ n bất kỳ nếu tổng thể có phân phối chuẩn

Độ lệch chuẩn của Số trung bình mẫu $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$ *được gọi là sai số chuẩn (Standard error).*

- Trường hợp tổng thể hữu hạn: Các điều kiện của định lý giới hạn trung tâm vẫn đúng. *Số trung bình mẫu vẫn có phân phối chuẩn nhưng phương sai được điều chỉnh hữu hạn.*

$$\overline{x} \sim N(\mu, \frac{\sigma^2}{n}\frac{N-n}{N-1})$$

Với điều kiện:

+ $n \geq 30$ nếu tổng thể có phân phối bất kỳ
+ $n \geq 15$ nếu tổng thể có phân phối đối xứng
+ n bất kỳ nếu tổng thể có phân phối chuẩn

Độ lệch chuẩn của Số trung bình mẫu $\sigma_{\overline{x}} = \sqrt{\frac{\sigma^2}{n}\frac{N-n}{N-1}}$ *được gọi là sai số chuẩn (Standard error).*

Thành phần $\frac{N-n}{N-1}$ được gọi là nhân tố điều chỉnh tổng thể hữu hạn (finite population correction factor).

Chú ý:

(1) Trong những tình huống thực hành, khi $n/N \leq 5$ (%) nhân tố điều chỉnh hữu hạn rất gần 1. Do đó, có thể sử dụng sai số chuẩn của mẫu ngẫu nhiên đơn giản từ tổng thể vô hạn thay cho mẫu ngẫu nhiên từ tổng thể hữu hạn.

(2) Trừ những trường hợp đặc biệt được lưu ý riêng, trong suốt giáo trình này sẽ giả định $n/N \leq 5$ (%) và $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$ được dùng để tính sai số chuẩn cho cả hai trường hợp.

4.3.2. Phân phối lấy mẫu của Tỉ lệ mẫu

Phân phối lấy mẫu của tỉ lệ mẫu là phân phối xác suất của tất cả các giá trị có khả năng của tỉ lệ mẫu.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

- Trường hợp tổng thể vô hạn:

Ta có, tỉ lệ mẫu $\overline{p} = \frac{x}{n}$. Trong đó x là số đơn vị tổng thể có biểu hiện nghiên cứu trên mẫu. Xác suất xuất hiện một đơn vị tổng thể có biểu hiện nghiên cứu trên mẫu bằng p không đổi. Do đó $n\overline{p} = x$ có phân phối nhị thức với kỳ vọng toán là np và phương sai là np(1-p). Phân phối nhị thức này xấp xỉ phân phối chuẩn khi np≥5 và n(1-p)≥5.

Suy ra, *Tỉ lệ mẫu* $\overline{p}$ *có phân phối chuẩn:*

$$\overline{p} \sim N\left[p, \frac{p(1-p)}{n}\right]$$ *Với điều kiện* np≥5 và n(1-p)≥5

Độ lệch chuẩn của Tỉ lệ mẫu $\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}}$ *được gọi là sai số chuẩn (Standard error).*

- Trường hợp tổng thể hữu hạn: Tương tự, tỉ lệ mẫu cũng có phân phối chuẩn.

$$\overline{p} \sim N\left[p, \frac{p(1-p)}{n}\frac{N-n}{N-1}\right]$$ *Với điều kiện* np≥5 và n(1-p)≥5

Độ lệch chuẩn của Tỉ lệ mẫu $\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}\frac{N-n}{N-1}}$ *được gọi là sai số chuẩn (Standard error).*

Thành phần $\frac{N-n}{N-1}$ được gọi là nhân tố điều chỉnh tổng thể hữu hạn (finite population correction factor).

Chú ý:

(1) Trong những tình huống thực hành, khi n/N ≤ 5 (%) nhân tố điều chỉnh hữu hạn rất gần 1. Do đó, có thể sử dụng sai số chuẩn của mẫu ngẫu nhiên đơn giản từ tổng thể vô hạn thay cho mẫu ngẫu nhiên từ tổng thể hữu hạn.

(2) Trừ những trường hợp đặc biệt được lưu ý riêng, trong suốt giáo trình này sẽ giả định n/N ≤ 5 (%) và $\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}}$ được dùng để tính sai số chuẩn cho cả hai trường hợp.

4.3.3. Phân phối lấy mẫu của phương sai mẫu

Khi tổng thể có phân phối chuẩn, theo phân phối Khi bình phương:

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{\sigma^2} \sim \chi^2_{n-1}$$

4.4. Ước lượng điểm

4.4.1. Các tiêu chuẩn ước lượng điểm

Ước lượng điểm là xác định một trị số gần đúng nhất của một tham số tổng thể dựa trên thống kê mẫu. Để bảo đảm độ tin cậy của ước lượng điểm, có ba tiêu chuẩn ước lượng điểm được đưa ra như sau.

- *Tiêu chuẩn không chệch:* Một thống kê $\hat{\theta}$ trên mẫu được gọi là ước lượng không chệch của tham số θ trên tổng thể, nếu: $E(\hat{\theta}) = \theta$.

- *Tiêu chuẩn vững:* Một thống kê $\hat{\theta}$ trên mẫu được gọi là ước lượng vững của tham số θ trên tổng thể, nếu $\hat{\theta}$ hội tụ theo xác suất về θ khi kích thước mẫu n lớn đến vô cùng.

Tức là: $\begin{cases} E(\hat{\theta}) = \theta \\ \lim_{n\to\infty} V(\hat{\theta}) = 0 \end{cases}$

- *Tiêu chuẩn hiệu quả:* Một thống kê $\hat{\theta}$ trên mẫu được gọi là ước lượng hiệu quả của tham số θ trên tổng thể, nếu nó là ước lượng không chệch và có phương sai nhỏ nhất so với mọi ước lượng không chệch khác được xây dựng trên cùng mẫu tổng quát đó. Tức là: V($\hat{\theta}$) = min.

4.4.2. Ước lượng điểm:

a. Ước lượng điểm của Số trung bình tổng thể:

Số trung bình mẫu có phân phối chuẩn: $\overline{x} \sim N(\mu, \frac{\sigma^2}{n})$

Suy ra: $E(\overline{x}) = \mu$ và $\lim_{n\to\infty} \sigma^2_{\overline{x}} = 0$

Vậy $\overline{x}$ là ước lượng không chệch và vững của μ.

Mặt khác, $\overline{x}$ có phân phối chuẩn nên theo bất đẳng thức Cramer-Rao, $\overline{x}$ là ước lượng hiệu quả của μ.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Do đó, ước lượng điểm tốt nhất của Số trung bình tổng thể μ là số trung bình mẫu $\bar{x}$.

b. Ước lượng điểm của Tỉ lệ tổng thể:

Lập luận tương tự trên đây, *ước lượng điểm tốt nhất của Tỉ lệ tổng thể P là tỉ lệ mẫu* $\bar{p}$.

c. Ước lượng điểm của Phương sai tổng thể:

Khi tổng thể có phân phối chuẩn, theo phân phối phương sai mẫu:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Theo phân phối Khi bình phương: $E(\frac{(n-1)S^2}{\sigma^2}) = E(\chi^2_{n-1}) = n-1$

Suy ra: $E(s^2) = \sigma^2$

Điều này cũng được thống kê toán chứng minh là đúng cho cả trường hợp tổng thể có phân phối bất kỳ.

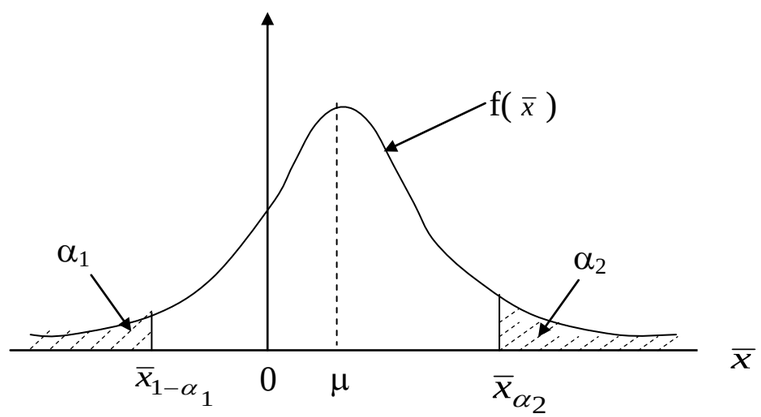
Do đó, ước lượng điểm của Phương sai tổng thể σ^2 *là phương sai mẫu:* s^2.

4.5. Ước lượng khoảng

Ước lượng khoảng là xác định một khoảng số thực sao cho xác suất để một tham số tổng thể rơi vào khoảng đó tương đối lớn.

4.5.1. Ước lượng khoảng của Số trung bình tổng thể

Theo định lý giới hạn trung tâm cần lấy mẫu kích thước n ≥30. Nếu muốn lấy mẫu kích thước n < 30, tổng thể phải có phân phối xấp xỉ chuẩn. Như đã biết, Số trung bình mẫu có phân phối chuẩn: $\bar{x} \sim N\left(\mu, \sigma^2_{\bar{X}}\right)$.



Trên đồ thị phân phối của Số trung bình mẫu, khoảng ước lượng với độ tin cậy 1- ($\alpha_1+\alpha_2$) của μ là khoảng $(\bar{x}_{1-\alpha_1}, \bar{x}_{\alpha_2})$.

a. Trường hợp biết trước phương sai tổng thể σ^2

Để tiện tra cứu các điểm phân vị, chuyển phân phối chuẩn của số trung bình mẫu về phân phối chuẩn tắc Z.

$$\bar{x} \sim N\left(\mu, \sigma^2_{\bar{x}}\right) \Rightarrow Z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \sim N(0,1) \Rightarrow \bar{x} = \mu + Z\sigma_{\bar{x}}$$

$$\Rightarrow \begin{cases} \bar{x}_{\alpha_2} = \mu + z_{\alpha_2}\sigma_{\bar{x}} \\ \bar{x}_{1-\alpha_1} = \mu + z_{1-\alpha_1}\sigma_{\bar{x}} = \mu - z_{\alpha_1}\sigma_{\bar{x}} \end{cases}$$

Vì vậy, khoảng ước lượng $(\bar{x}_{1-\alpha_1}, \bar{x}_{\alpha_2})$ có thể viết lại là:

$$\left(\mu - z_{\alpha_1}\sigma_{\bar{x}},\ \mu + z_{\alpha_2}\sigma_{\bar{x}}\right)$$

Khi ước lượng với mẫu cụ thể, ta thay μ bằng ước lượng điểm tốt nhất của nó là $\bar{x}$. Mặt khác, trong thực tế, người ta thường dùng khoảng ước lượng đối xứng với $\alpha_1 = \alpha_2 = \alpha/2$.

Do đó, khoảng tin cậy của μ*:* $(\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$

Trong đó: * $\bar{x}$ *: Số trung bình mẫu*

* $z_{\alpha/2}$*: Phân vị chuẩn mức* $\alpha/2$

Chú ý:

- Đại lượng $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ *được gọi là biên sai số (margin of error) hay phạm vi sai số lấy mẫu*

- Đại lượng $D = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ *được gọi là độ dài khoảng tin cậy*

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

b. Trường hợp chưa biết phương sai tổng thể σ^2

Do chưa biết phương sai tổng thể, có thể chuyển phân phối chuẩn của trung bình mẫu về phân phối Student (xem phân phối Student).

$$\bar{x} \sim N\left(\mu, \sigma_{\bar{X}}^2\right) \Rightarrow \frac{\bar{x}-\mu}{\sqrt{\frac{s^2}{n}}} \sim T_{n-1} \Rightarrow \bar{x} = \mu + T_{n-1}\sqrt{\frac{s^2}{n}}$$

Suy ra: $\bar{x}_{1-\alpha_1} = \mu - t_{n-1,\alpha_1}\sqrt{\frac{s^2}{n}}$

Và: $\bar{x}_{\alpha_2} = \mu + t_{n-1,\alpha_2}\sqrt{\frac{s^2}{n}}$

Do đó, khoảng tin cậy của μ là: $(\bar{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}})$

Trong đó, $t_{n-1,\alpha/2}$ là phân vị Student với n-1 bậc tự do, mức $\alpha/2$

Ví dụ: Để kiểm tra chất lượng một lô hàng lớn, người ta chọn ngẫu nhiên đơn giản 100 sản phẩm. Kết quả thu được như trang sau.

Trọng lượng (kg) (x_i)	Số sản phẩm (f_i)
4,8	10
4,9	20
5,0	30
5,1	20
5,2	20

Ta ước lượng trọng lượng trung bình một sản phẩm của lô hàng như sau:

Ước lượng điểm: $\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = 5{,}02$ *(kg)*

Khoảng tin cậy 95%: $(\bar{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}})$

Trong đó: $s = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{n-1}} = \sqrt{0{,}0162} = 0{,}127$

Tra bảng Student: $t_{n-1,\,\alpha/2} = t_{99,\,0{,}025} = 1{,}99$

Thay số: $(5{,}02 \pm 1{,}99\frac{0{,}127}{\sqrt{100}})$

Khoảng tin cậy: (4,995, 5,065) kg

4.5.2. Ước lượng khoảng của Tỉ lệ tổng thể

Với điều kiện np≥5 và n(1-p)≥5, Tỉ lệ mẫu có phân phối chuẩn :

$$\bar{p} \sim N\left[p, \frac{p(1-p)}{n}\right]$$

Do đó, lập luận tương tự trường hợp 4.5.1, ta có:

Khoảng tin cậy của Tỉ lệ tổng thể p là: $(\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}})$

Trong đó: * $\bar{p}$ *: Tỉ lệ mẫu*

* $z_{\alpha/2}$*: Phân vị chuẩn mức $\alpha/2$*

Chú ý:

- Đại lượng $z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ *được gọi là biên sai số (margin of error)*

- Đại lượng $D = 2z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ *được gọi là độ dài khoảng tin cậy*

Ví dụ: Để kiểm tra chất lượng của một dây chuyền sản xuất, người ta chọn 1000 sản phẩm ngẫu nhiên, cứ một phút một sản phẩm vừa ra khỏi dây chuyền. Kết quả cho thấy, có 20 sản phẩm có khuyết tật. Ta ước lượng Tỉ lệ sản phẩm có khuyết tật như sau:

Ước lượng điểm: $\bar{p} = \frac{20}{1000} = 0{,}02 = 2\%$

Khoảng tin cậy 95%: $(\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}})$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Tra bảng phân vị chuẩn: $z_{\alpha/2} = z_{0,025} = 1,96$

Thay số: $(0,02 \pm 1,96\sqrt{\frac{0,02(1-0,02)}{1000}})$

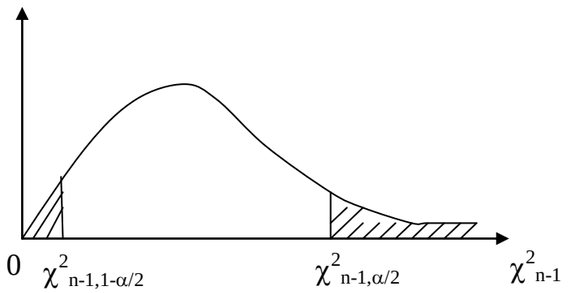
Hay: (0,0113, 0,0287)

Hoặc: (1,13, 2,87) %

4.5.3. Ước lượng khoảng của phương sai tổng thể

Khi tổng thể có phân phối chuẩn, theo phân phối phương sai mẫu:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$



Do đó: $P\left(\chi^2_{n-1,1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1,\alpha/2}\right) = 1-\alpha$

Suy ra: $P\left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right) = 1-\alpha$

Khoảng tin cậy của phương sai tổng thể σ^2 là:

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right)$$

Ví dụ: Để kiểm tra chất lượng một lô hàng lớn, người ta chọn ngẫu nhiên đơn giản 100 sản phẩm. Kết quả thu được như trang sau.

Trọng lượng (kg) (x_i)	Số sản phẩm (f_i)
4,8	10
4,9	20
5,0	30
5,1	20
5,2	20

Khoảng tin cậy 95% của phương sai trọng lượng lô hàng:

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right)$$

Trong đó: $\bar{x} = 5,02\ (kg)$, $s^2 = 0,0162$

Tra bảng Khi bình phương: $\chi^2_{n-1,\alpha/2} = \chi^2_{99;\,0,025} = 128,4$

$$\chi^2_{n-1,1-\alpha/2} = \chi^2_{99;\,0,975} = 73,4$$

Thay số: $\left(\frac{(100-1)0,0162}{128,4}, \frac{(100-1)0,0162}{73,4}\right)$

Hay: (0,0125; 0,0219)

4.6. Xác định kích thước mẫu

Câu hỏi đặt ra là kích thước mẫu lớn cỡ nào thì sẽ thu được ước lượng của một tham số tổng thể đạt yêu cầu về độ chính xác. Câu trả lời tùy thuộc vào loại tham số cần ước lượng.

4.6.1. Kích thước mẫu khi ước lượng số trung bình

Đặt sai số biên mong muốn là E. Từ công thức tính sai số biên, ta có:

$$E = z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}$$

Suy ra: $$n = \frac{z^2_{\alpha/2}}{E^2}\sigma^2$$

Chú ý:

** Độ tin cậy 1-α thường được chọn từ 90 đến 99%. Giá trị thường dùng nhất là 95%.*

** Trong thực hành thường không có sẵn Phương sai tổng thể, người ta thường chọn một trong những giải pháp sau:*

+ Lấy phương sai lớn nhất trong các cuộc điều tra trước (nếu có)

+ Lấy phương sai hiện tượng tương tự ở nơi khác (nếu có)

+ Lấy $\sigma = (x_{max}-x_{min})/4$ với x_{max} và x_{min} là lượng biến lớn nhất và nhỏ nhất quan sát hay phán đoán được trong thực tế.

+ Điều tra trên phạm vi nhỏ để ước tính $\sigma^2 \approx s^2$.

Ví dụ 1: Xét lô hàng ở ví dụ mục 4.5.1. Nếu đặt ra độ tin cậy 99%, sai số biên mong muốn là E =0,01kg , kích thước mẫu sẽ là:

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

$$n = \frac{z_{\alpha/2}^2}{E^2}\sigma^2$$

Trên mẫu $n^* = 100$ sản phẩm ở ví dụ trước, ta có $s^2 = 0{,}0162$.

Tra bảng phân vị chuẩn: $z_{\alpha/2} = z_{0{,}005} = 2{,}575$

Do đó: $$n = \frac{2{,}575^2}{0{,}01^2} \cdot 0{,}0162 = 1075 SP$$

Lưu ý: Kích thước mẫu trên có thể quá lớn, chi phí kiểm tra sẽ là quá cao. Doanh nghiệp có thể không chấp nhận được. Để giải quyết tình trạng trên, có thể mở rộng sai số biên mong muốn lên E =0,02 và giảm độ tin cậy xuống 95%. Kích thước mẫu sẽ là:

$$n = \frac{1{,}96^2}{0{,}02^2} \cdot 0{,}0162 = 156 SP$$

Kích thước mẫu này có thể chấp nhận được.

Ví dụ 2: Cần phải kiểm tra mẫu ngẫu nhiên bao nhiêu sản phẩm để ước lượng trọng lượng trung bình một sản phẩm của một lô hàng với độ tin cậy 95% và độ dài khoảng tin cậy 0,05kg. Theo ước đoán, sản phẩm nặng nhất của lô hàng là 5,2kg, sản phẩm nhẹ nhất của lô hàng là 4,8 kg.

$\sigma \approx (x_{max} - x_{min})/4 = (5{,}2 - 4{,}8)/4 = 0{,}1$

Kích thước mẫu: $$n = \frac{z_{\alpha/2}^2}{E^2}\sigma^2$$

$$n = 4 \cdot \frac{1{,}96^2}{0{,}025^2} \cdot 0{,}1^2 = 62 SP$$

4.6.2. Kích thước mẫu khi ước lượng các Tỉ lệ

Đặt sai số biên mong muốn là E. Từ công thức tính sai số biên, ta có:

$$E = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

Suy ra: $$n = \frac{z_{\alpha/2}^2}{E^2} p(1-p)$$

Chú ý: *Trong thực hành thường không có sẵn phương sai tỉ lệ tổng thể, người ta thường chọn một trong những cách sau để thay thế:*

+ Lấy phương sai tỉ lệ lớn nhất trong các cuộc điều tra trước (nếu có)
+ Lấy phương sai tỉ lệ hiện tượng tương tự ở nơi khác (nếu có)
+ Điều tra trên phạm vi nhỏ để ước tính.
+ Lấy p=0,5.

Ví dụ 1: Cần phải kiểm tra mẫu ngẫu nhiên bao nhiêu sản phẩm để ước lượng tỉ lệ sản phẩm trên 5kg của một lô hàng với độ tin cậy 99% và sai số biên mong muốn là 0,1.

Một mẫu ngẫu nhiên tạm thời 30 SP được chọn.

Trọng lượng (kg) (x_i)	Số sản phẩm (n_i)
4,8	2
4,9	5
5,0	15
5,1	4
5,2	4

Trên mẫu tạm thời: $\overline{p}^* = 8/30 = 0{,}267$

Kích thước mẫu: $$n = \frac{z_{\alpha/2}^2}{E^2} p(1-p)$$

$$n = \frac{2{,}575^2}{0{,}1^2} 0{,}267(1 - 0{,}267) = 130\ (SP)$$

Ví dụ 2: Cần phải kiểm tra mẫu ngẫu nhiên bao nhiêu sản phẩm để ước lượng tỉ lệ sản phẩm có trọng lượng dưới 2kg của một lô hàng lớn với độ tin cậy 99% và sai số biên mong muốn là 0,1.

Không có thông tin nào về tỉ lệ tổng thể, chọn p = 0,5.

Kích thước mẫu: $$n = \frac{z_{\alpha/2}^2}{E^2} p(1-p)$$

$$n = \frac{2{,}575^2}{0{,}1^2} . 0{,}5(1 - 0{,}5) = 166\ (SP)$$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Chương 5

KIỂM ĐỊNH THAM SỐ VÀ ƯỚC LƯỢNG SỰ KHÁC BIỆT

Trong hoạt động quản lý kinh tế xã hội, chúng ta thường đối mặt với các quyết định như: phải hủy bỏ một lô hàng vì nghi ngờ rằng nó có khả năng không bảo đảm vệ sinh an toàn thực phẩm, cấm đưa ra thị trường một loại thuốc để chữa bệnh vì lo sợ rằng nó có tác dụng phụ nguy hiểm. Các quyết định này có thể đúng hoặc sai. Khi sự đúng hay sai này có ý nghĩa kinh tế xã hội lớn trên một tổng thể số lớn hiện có hoặc khuếch đại về mặt thời gian hay không gian thì việc xác minh sự đúng hay sai trước khi quyết định là rất quan trọng. *Việc xác minh các nhận định chủ quan trên cả tổng thể số lớn là rất tốn kém thời gian, tiền bạc và đôi khi không thể thực hiện được. Do đó, xác minh nó dựa trên bằng chứng mẫu là vô cùng cần thiết và hiệu quả.*

5.1. Giả thuyết và các sai lầm trong kiểm định giả thuyết thống kê

5.1.1. Giả thuyết thống kê

Giả thuyết thống kê là một nhận định (tentative assumption) nào đó được đưa ra về một hay nhiều tổng thể để trả lời một câu hỏi đặt ra nào đó. Chẳng hạn, giả thuyết cho rằng thu nhập trung bình của người dân thành phố Y là 3 triệu đồng hay giả thuyết cho rằng thu nhập và giới tính của người dân thành phố X có liên hệ với nhau… Một giả thuyết thống kê đầy đủ bao gồm giả thuyết không và giả thuyết đối. Giả thuyết không thường được ký hiệu là H_o và giả thuyết đối (ngược lại với giả thuyết không) thường được ký hiệu là H_1 hoặc H_a.

Ví dụ:

Giả thuyết về Số trung bình tổng thể với đối thuyết hai phía:

$H_o : \mu = \mu_o$ (μ_o hằng số)

$H_1 : \mu \neq \mu_o$

Giả thuyết về Số trung bình tổng thể với đối thuyết phía phải:

$H_o : \mu \leq \mu_o$ (μ_o hằng số)

$H_1 : \mu > \mu_o$

Giả thuyết về Số trung bình tổng thể với đối thuyết phía trái:

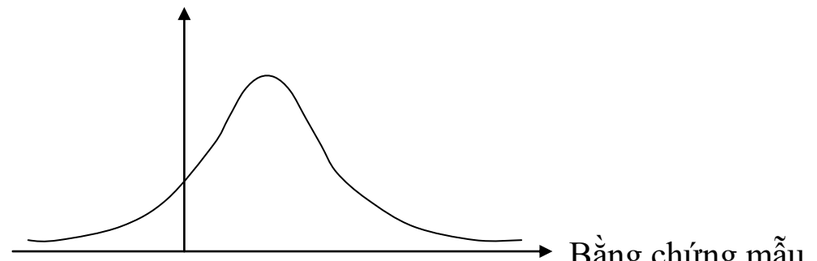
$H_o : \mu \geq \mu_o$ (μ_o hằng số)

$H_1 : \mu < \mu_o$

Để đánh giá một giả thuyết đưa ra là đúng hay sai trước khi đưa vào vận dụng ta cần phải tiến hành kiểm định giả thuyết ấy.

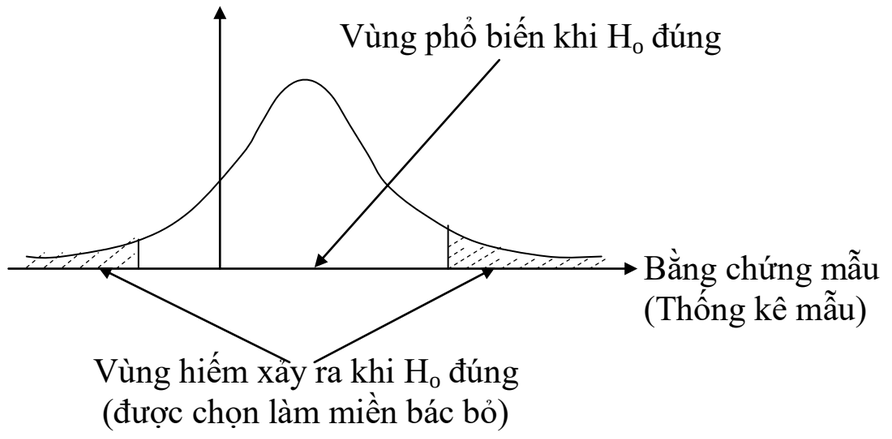
Nguyên lý chung của kiểm định giả thuyết:

- Trước hết cần xây dựng cặp giả thuyết không-giả thuyết đối một cách chính xác, trả lời đúng câu hỏi đặt ra về hiện tượng nghiên cứu.
- Lấy mẫu ngẫu nhiên từ tổng thể, thu thập dữ liệu làm bằng chứng để so sánh với giả thuyết.
- Việc lấy mẫu ngẫu nhiên sẽ cho bằng chứng mẫu mang tính ngẫu nhiên rất khác nhau giữa các lần lấy mẫu khác nhau. Vì vậy, cần xác định được phân phối mẫu cho mọi bằng chứng mẫu trong điều kiện giả định giả thuyết H_o đúng.



Phân phối mẫu khi H_o đúng

- Xác định vùng phân phối hiếm xảy ra của bằng chứng mẫu khi H_o đúng (nhưng sẽ là vùng phổ biến của bằng chứng mẫu khi H_o sai). Đây là miền nghi ngờ cao cho H_o đúng và được chọn làm miền bác bỏ của H_o. Mức xác suất miền bác bỏ của phân phối mẫu thường được chọn ở mức nhỏ khoảng 5% hay 1%.



- Nếu bằng chứng mẫu kiểm định rơi vào miền bác bỏ, kết luận bác bỏ H_o với mức xác suất sai lầm có thể gặp phải của miền bác bỏ. Nếu bằng chứng mẫu kiểm định không rơi vào miền bác bỏ, chưa đủ cơ sở bác bỏ H_o.

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

5.1.2. Các loại sai lầm và độ giá trị của kiểm định giả thuyết

Có hai loại sai lầm có khả năng xuất hiện trong kiểm định giả thuyết thống kê là sai lầm loại I và sai lầm loại II.

- Sai lầm loại I là sai lầm khi giả thuyết H_o đúng mà ta bác bỏ giả thuyết này.

$$P(\text{sai lầm loại I}) = P(\text{bác bỏ } H_o / H_o \text{ đúng}) = \alpha$$

- Sai lầm loại II là sai lầm khi giả thuyết H_o sai mà ta chấp nhận giả thuyết này.

$$P(\text{sai lầm loại II}) = P(\text{chấp nhận } H_o / H_o \text{ sai}) = \beta$$

Khi kiểm định giả thuyết, ta không biết giả thuyết H_o là đúng hay sai nên có khả năng mắc một trong hai loại sai lầm trên. Càng dễ chấp nhận H_o càng giảm khả năng mắc sai lầm loại I nhưng càng tăng khả năng mắc sai lầm loại II, và ngược lại.

Ngoài ra, khi kiểm định người ta còn quan tâm đến độ giá trị của kiểm định (Power) đó là xác suất bác bỏ đúng H_o khi H_o sai.

$$\text{Độ giá trị} = P(\text{bác bỏ } H_o | H_o \text{ sai}) = 1 - \beta$$

Độ giá trị của kiểm định phụ thuộc thuận chiều vào 3 yếu tố: khả năng đúng của H_1, mức ý nghĩa α lựa chọn, và kích thước mẫu.

Trong điều kiện hạn chế về kích thước mẫu vì lý do chi phí và yêu cầu kiểm soát sai lầm loại I, khả năng tăng độ giá trị dựa chủ yếu vào việc nâng cao khả năng đúng của H_1. *Vì vậy, trong xây dựng giả thuyết, H_1 có tầm quan trọng đặc biệt trong việc nâng cao độ giá trị của kiểm định.*

5.1.3. Các bước kiểm định giả thuyết thống kê

Theo nguyên lý kiểm định nói trên, các bước kiểm định giả thuyết thống kê bao gồm:

- Xây dựng cặp giả thuyết không và giả thuyết đối;

- Xác định độ ý nghĩa của kiểm định (xác suất miền bác bỏ, xác suất sai lầm loại 1 cho phép)

- Thu thập dữ liệu mẫu và tính tiêu chuẩn kiểm định;

- Tra bảng điểm tới hạn ứng với mức ý nghĩa α của phân phối mẫu (theo tiếp cận điểm tới hạn) hoặc tra bảng giá trị xác suất P của phân phối mẫu ứng với tiêu chuẩn kiểm định (theo tiếp cận giá trị P);

- So sánh tiêu chuẩn kiểm định với điểm tới hạn (hoặc giá trị xác suất P với mức ý nghĩa α) và kết luận có nên bác bỏ H_o.

5.1.4 Xây dựng giả thuyết và kiểm soát các loại sai lầm

Theo nguyên lý kiểm định ở trên, xác suất miền bác bỏ trên phân phối bằng chứng mẫu (với giả định H_o đúng) chính là xác suất có thể phạm sai lầm loại I khi bác bỏ H_o nếu bằng chứng mẫu rơi vào miền này. Vì vậy, *sai lầm loại I được kiểm soát trực tiếp trong thủ tục kiểm định* bằng xác suất (α) của miền bác bỏ. Mức xác suất (α) này thường được gọi là mức ý nghĩa của kiểm định.

Tuy nhiên, *sai lầm loại II và độ giá trị kiểm định không được kiểm soát trực tiếp trong thủ tục kiểm định.*

Để kiểm soát sai lầm và độ giá trị nói chung, có ba kiểu đặt giả thuyết thường được sử dụng là: kiểu giả thuyết nghiên cứu, kiểu giả thuyết phòng ngừa, và kiểu giả thuyết kiểm tra để ra các quyết định thông thường.

- Kiểu giả thuyết nghiên cứu: Giả thuyết đối (H_1) được xây dựng trước một cách có cơ sở để bảo đảm xác suất đúng cao. Cơ sở để xây dựng H_1 thường là các lý thuyết, các kết quả đã kiểm định trước đó, nhận định của các nhà khoa học khác. Giả thuyết không (H_o) chỉ cần đặt ngược lại H_1.

Trong các nghiên cứu khoa học, giả thuyết nghiên cứu được thể hiện trong H_1. Việc bác bỏ H_o thừa nhận giả thuyết nghiên cứu (H_1) chỉ phạm sai lầm loại I khoảng 5% hay 1%. Sai lầm loại II và độ giá trị được kiểm soát một cách gián tiếp thông qua cơ sở xây dựng H_1.

- Kiểu giả thuyết phòng ngừa: Giả thuyết không (H_o) luôn luôn được xác định theo kiểu phòng ngừa: không an toàn, không tốt,... Giả thuyết đối (H_1) được xác định ngược lại theo kiểu: an toàn, tốt,...

Giả thuyết kiểu này chỉ nhằm kiểm soát trực tiếp sai lầm loại I trong thủ tục kiểm định ở mức xác suất (α) rất thấp. Sai lầm loại II sẽ là thứ yếu không cần kiểm soát. Kiểu giả thuyết này chỉ nhằm bảo vệ một cách thiên vị một đối tượng nào đó.

Ví dụ giả thuyết về an toàn thực phẩm, an toàn dược phẩm nhằm bảo vệ người tiêu dùng, giả thuyết về các quyết định có liên quan đến sự sống còn của một doanh nghiệp.

Chẳng hạn, giả thuyết về an toàn thực phẩm thường được xây dựng kiểu phòng ngừa: H_o cho rằng lô thực phẩm là không an toàn mà không cần xét đoán điều gì cả. H_1 được đặt ngược lại H_o là không an toàn.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Giả thuyết:

$H_o : p \geq p_o$ (p là tỉ tệ dư lượng kháng sinh trong thực phẩm)

$H_1 : p < p_o$ (p_o là tỉ lệ dư lượng kháng sinh tối đa được phép)

- *Kiểu giả thuyết kiểm tra để ra các quyết định thông thường*: Đây là trường hợp người ra quyết định phải lựa chọn một trong hai loại hành động có tầm quan trọng gần như nhau. Một hành động gắn với giả thuyết không và một hành động gắn với giả thuyết đối.

Căn cứ xây dựng giả thuyết trong trường hợp này là các cam kết, các tiêu chuẩn đặt ra, các nhận định của những người có trách nhiệm nào đó... Để phản ánh đúng tình huống cần kiểm tra, H_o hoặc H_1 có thể được xây dựng trước. Giả thuyết kia chỉ cần xây dựng ngược lại. Cách đặt giả thuyết này cũng chỉ kiểm soát trực tiếp được sai lầm loại I với xác suất (α) đặt ra. Sai lầm loại II rất khó kiểm soát.

Trường hợp người ra quyết định tự xây dựng giả thuyết, phải dựa trên sự quan sát kỹ lưỡng các dấu hiệu trên thực tế để xây dựng H_1 trước một cách có cơ sở và ưu tiên xác suất đúng của nó để giúp kiểm soát gián tiếp sai lầm loại II. Giả thuyết phải được xây dựng độc lập với dữ liệu mẫu được dùng để kiểm định. Nếu giả thuyết được xây dựng dựa trên quan sát dữ liệu mẫu dùng để kiểm định thì kiểm định đó không có ý nghĩa.

Chú ý: Dù sử dụng cách xây dựng giả tthuyết nào trên đây, theo nguyên lý kiểm định ở trên, để phân phối của bằng chứng mẫu xác định thì *H_o phải là một giả thuyết xác định.* Tức là, H_o chỉ là một khả năng duy nhất trong nhiều khả năng của giả thuyết. Cụ thể:

- *Với giả thuyết so sánh:* H_o phải là so sánh bằng. Đây là khả năng duy nhất trong các mức so sánh. So sánh bằng phải có dấu bằng ở các dạng "=" hoặc "≥" hoặc "≤".

- *Với giả thuyết về liên hệ:* H_o phải là không có liên hệ. Đây là khả năng duy nhất trong các mức liên hệ khác nhau.

- *Với giả thuyết về phân phối:* H_o phải có phân phối được chỉ định. Đây là khả năng duy nhất trong các kiểu phân phối có khả năng.

5.2. Kiểm định giả thuyết về tham số một tổng thể

Kiểm định tham số là loại kiểm định các giả thuyết về các tham số tổng thể dựa trực tiếp trên dữ liệu của mẫu. Dữ liệu này phải được đo trên các thang đo có độ chặt chẽ cao là thang đo tỉ lệ hoặc thang đo khoảng và đáp ứng các điều kiện về phân phối đặt ra trong các thủ tục kiểm định.

5.2.1. Kiểm định giả thuyết về Số trung bình một tổng thể

Giả thuyết về Số trung bình tổng thể với đối thuyết hai phía:

$H_o : \mu = \mu_o$ (μ_o là một hằng số)

$H_1 : \mu \neq \mu_o$

Giả thuyết về Số trung bình tổng thể với đối thuyết phía phải:

$H_o : \mu \leq \mu_o$

$H_1 : \mu > \mu_o$

Giả thuyết về Số trung bình tổng thể với đối thuyết phía trái:

$H_o : \mu \geq \mu_o$

$H_1 : \mu < \mu_o$

Theo định lý giới hạn trung tâm cần lấy mẫu ngẫu nhiên kích thước n ≥30. Nếu muốn lấy mẫu ngẫu nhiên kích thước n < 30 thì tổng thể phải có phân phối đối xứng hoặc xấp xỉ chuẩn.

a. Trường hợp biết trước phương sai tổng thể σ^2

Số trung bình mẫu có phân phối chuẩn: $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

$$\Rightarrow \quad Z = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

Nếu H_o đúng, thì $\mu = \mu_o$: $$Z = \frac{\bar{x} - \mu_o}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

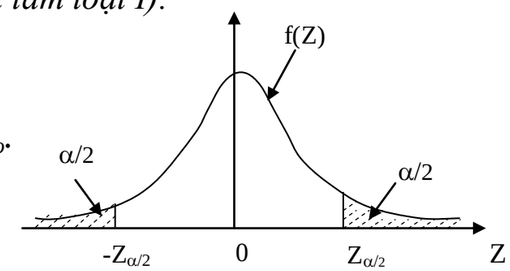
Do đó, *tiêu chuẩn kiểm định:* $$z = \frac{\bar{x} - \mu_o}{\sqrt{\frac{\sigma^2}{n}}}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):

+ Kiểm định hai phía:

* $|z| \geq z_{\alpha/2}$: *Bác bỏ H_o, chấp nhận H_1*

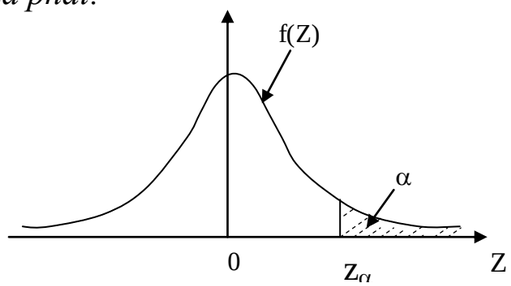
* $|z| < z_{\alpha/2}$ *: Chưa đủ cơ sở bác bỏ H_o.*



Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

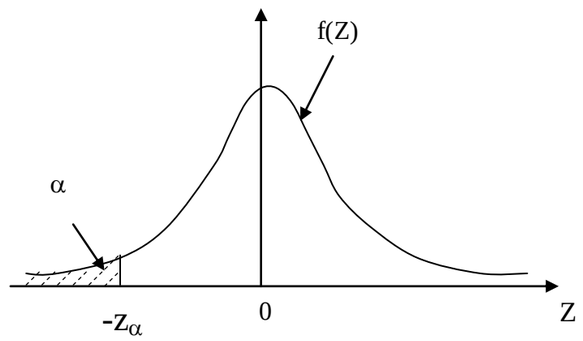
+ *Kiểm định phía phải:*



* $z \geq z_\alpha$: *Bác bỏ H_o, chấp nhận H_1*

* $z < z_\alpha$ *: Chưa đủ cơ sở bác bỏ H_o.*

+ *Kiểm định phía trái:*



* $z \leq -z_\alpha$: *Bác bỏ H_o, chấp nhận H_1*

* $z > -z_\alpha$ *: Chưa đủ cơ sở bác bỏ H_o.*

Chú ý: Kết luận kiểm định theo tiếp cận giá trị P như sau:

- *Tra bảng phân vị giá trị P của tiêu chuẩn kiểm định.*

* Kiểm định một phía: P-value = Sig (one-sided) = $P(Z \geq |z|)$

* Kiểm định hai phía: P-value = Sig (two-sided) = $2*P(Z \geq |z|)$

- *Kết luận dựa vào giá trị P:*

+ Kiểm định hai phía:

* Sig (two-sided) ≤ α : Bác bỏ H_o, chấp nhận H_1

* Sig (two-sided) > α : Chưa đủ cơ sở bác bỏ H_o.

+ Kiểm định phía phải:

* Sig (one-sided) ≤ α và z > 0 : Bác bỏ H_o, chấp nhận H_1

* Sig (one-sided) > α hoặc z ≤ 0: Chưa đủ cơ sở bác bỏ H_o.

+ Kiểm định phía trái:

* Sig (one-sided) ≤ α và z < 0 : Bác bỏ H_o, chấp nhận H_1

* Sig (one-sided) > α hoặc z ≥ 0: Chưa đủ cơ sở bác bỏ H_o.

b. Trường hợp chưa biết phương sai tổng thể σ^2

Số trung bình mẫu có phân phối chuẩn: $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

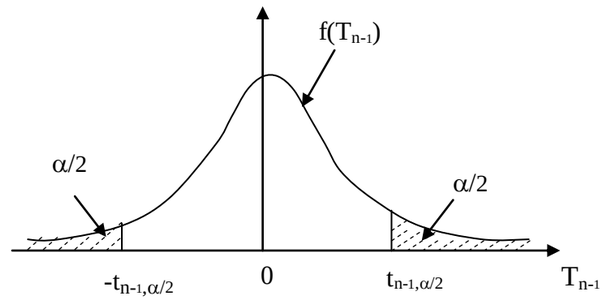
=> $\frac{\bar{x}-\mu}{\sqrt{\frac{s^2}{n}}} \sim T_{n-1}$ (xem phân phối Student).

Nếu H_o đúng, thì $\mu = \mu_o$: $\frac{\bar{x}-\mu_o}{\sqrt{\frac{s^2}{n}}} \sim T_{n-1}$

Do đó, *tiêu chuẩn kiểm định:* $t = \frac{\bar{x}-\mu_o}{\sqrt{\frac{s^2}{n}}}$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

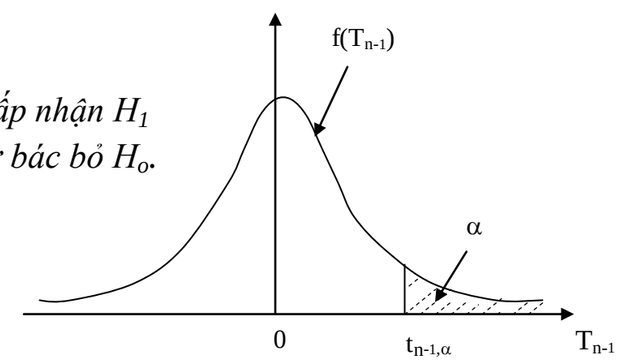
+ *Kiểm định hai phía:*



* $|t| \geq t_{n-1, \alpha/2}$: *Bác bỏ H_o, chấp nhận H_1*

* $|t| < t_{n-1, \alpha/2}$ *: Chưa đủ cơ sở bác bỏ H_o.*

+ *Kiểm định phía phải:*

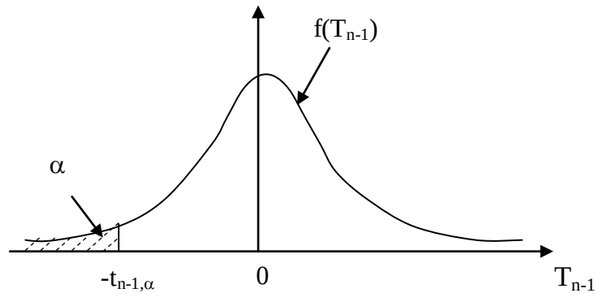


* $t \geq t_{n-1, \alpha}$: *Bác bỏ H_o, chấp nhận H_1*

* $t < t_{n-1, \alpha}$ *: Chưa đủ cơ sở bác bỏ H_o.*

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

+ *Kiểm định phía trái:*



* $t \le -t_{n-1,\alpha}$: *Bác bỏ H_o, chấp nhận H_1*

* $t > -t_{n-1,\alpha}$ *: Chưa đủ cơ sở bác bỏ H_o.*

Ví dụ: Tiêu chuẩn chất lượng đặt ra cho một loại sản phẩm của nhà máy là trọng lượng trung bình một sản phẩm đúng bằng 5kg. Để kiểm tra, mẫu 15 sản phẩm được chọn ngẫu nhiên. Kết quả thu được như sau:

Trọng lượng (kg) (x_i)	Số sản phẩm (n_i)
4,7	2
4,8	3
4,9	6
5,0	3
5,1	1

Biết rằng trọng lượng sản phẩm tuân theo phân phối chuẩn. Kiểm định với $\alpha = 0{,}01$.

Giả thuyết: $H_o : \mu = \mu_o$ (với: $\mu_o = 5$ kg)

$H_1 : \mu \neq \mu_o$

Tiêu chuẩn kiểm định: $$t = \frac{\bar{x} - \mu_o}{\sqrt{\frac{s^2}{n}}}$$

Trong đó: $$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = 4{,}887$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{n-1} = 0{,}0126$$

=> $$t = \frac{\bar{x} - \mu_o}{\sqrt{\frac{s^2}{n}}} = \frac{4{,}887 - 5}{\sqrt{\frac{0{,}0126}{15}}} = -3{,}9$$

Tra bảng: $t_{n-1,\,\alpha/2} = t_{14;\,0{,}005} = 2{,}98$

$|t| \ge t_{n-1,\,\alpha/2}$ => Bác bỏ H_o, chấp nhận H_1

Chú ý: Kết luận theo tiếp cận giá trị P như sau:

- Tra hàm T.DIST.2T trong Exell với 14 bậc tự do của giá trị |t| =3,9:

P-value = Sig (two-sided) = $2*P(T_{n-1} \ge |t|) = 2*P(T_{n-1} \ge 3{,}9) = 0{,}0032$

- Vì Sig (two-sided) < α. Bác bỏ H_o, chấp nhận H_1 với xác suất sai lầm loại một là 0,0032.

5.2.2. Kiểm định giả thuyết về Tỉ lệ một tổng thể

Giả thuyết về Tỉ lệ một tổng thể với đối thuyết hai phía:

$H_o : p = p_o$ (p_o là một hằng số)

$H_1 : p \neq p_o$

Giả thuyết về Tỉ lệ một tổng thể với đối thuyết phía phải:

$H_o : p \le p_o$

$H_1 : p > p_o$

Giả thuyết về Tỉ lệ một tổng thể với đối thuyết phía trái:

$H_o : p \ge p_o$

$H_1 : p < p_o$

Theo điều kiện của phân phối lấy mẫu của tỉ lệ mẫu, ta cần lấy mẫu ngẫu nhiên kích thước n đạt điều kiện $np_o \ge 5$ và $n(1-p_o) \ge 5$ để kiểm định. Khi đó, tỉ lệ mẫu có phân phối chuẩn.

$$\bar{p} \sim N\left[p, \frac{p(1-p)}{n}\right]$$

Chuyển về phân phối chuẩn tắc: $$Z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

Nếu H_o đúng, thì $p = p_o$: $$Z = \frac{\bar{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} \sim N(0,1)$$

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Do đó, *tiêu chuẩn kiểm định:* $$z = \frac{\bar{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:

* $|z| \geq z_{\alpha/2}$: *Bác bỏ H_o, chấp nhận H_1*
* $|z| < z_{\alpha/2}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía phải:

* $z \geq z_{\alpha}$: *Bác bỏ H_o, chấp nhận H_1*
* $z < z_{\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía trái:

* $z \leq - z_{\alpha}$: *Bác bỏ H_o, chấp nhận H_1*
* $z > - z_{\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

Ví dụ: Một công nhân cho rằng tỉ lệ sản phẩm hỏng của chiếc máy hiện nay đã không còn nhỏ hơn 0,1% như tiêu chuẩn chất lượng đặt ra. Để kiểm tra, người công nhân sản xuất thử 2000 sản phẩm thấy có 1 sản phẩm hỏng. Kiểm định với $\alpha = 0{,}01$.

Giả thuyết: $H_o : p \geq p_o$ (với $p_o = 0{,}001$)

$H_1 : p < p_o$

Tiêu chuẩn kiểm định: $$z = \frac{\bar{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$$

Trong đó: $$\bar{p} = \frac{1}{2000} = 0{,}0005$$

$$z = \frac{\bar{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} = \frac{0{,}0005 - 0{,}001}{\sqrt{\frac{0{,}001(1-0{,}001)}{2000}}} = -0{,}71$$

Tra bảng: $- z_{\alpha} = - z_{0{,}01} = - 2{,}328$

* $z > - z_{\alpha}$: Chưa đủ cơ sở bác bỏ H_o (có thể tạm chấp nhận H_o với sai lầm loại II nào đó).

Chú ý: Kết luận theo tiếp cận giá trị P như sau:

- Tra hàm NORM.S.DIST trong Exell của giá trị |z| =0,71 để xác định:

P-value = Sig (one-sided) = $P(Z \geq |z|) = P(Z \geq 0{,}71) = 0{,}239$

- Vì Sig (one-sided) > α: Chưa đủ cơ sở bác bỏ H_o.

5.2.3. Kiểm định giả thuyết về Phương sai một tổng thể

Giả thuyết về Phương sai tổng thể với đối thuyết hai phía:

$H_o : \sigma^2 = \sigma^2_o$ (σ^2_o là một hằng số)

$H_1 : \sigma^2 \neq \sigma^2_o$

Giả thuyết về Phương sai tổng thể với đối thuyết phía phải:

$H_o : \sigma^2 \leq \sigma^2_o$

$H_1 : \sigma^2 > \sigma^2_o$

Giả thuyết về Phương sai tổng thể với đối thuyết phía trái:

$H_o : \sigma^2 \geq \sigma^2_o$

$H_1 : \sigma^2 < \sigma^2_o$

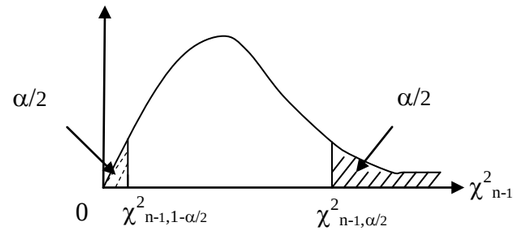
Để kiểm định, *lấy mẫu ngẫu nhiên kích thước n*. Nếu tổng thể có phân phối chuẩn, theo phân phối lấy mẫu của phương sai mẫu: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

Nếu H_o đúng, thì $\sigma^2 = \sigma^2_o$: $\frac{(n-1)S^2}{\sigma^2_o} \sim \chi^2_{n-1}$

Do đó, *tiêu chuẩn kiểm định:* $$\chi^2 = \frac{(n-1)s^2}{\sigma^2_o}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:

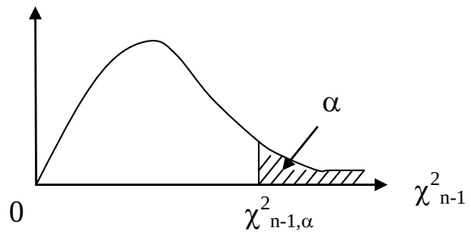


* $\chi^2 \leq \chi^2_{n-1,1-\alpha/2}$ *hoặc* $\chi^2 \geq \chi^2_{n-1,\alpha/2}$ *: Bác bỏ H_o, chấp nhận H_1*

* $\chi^2_{n-1,1-\alpha/2} < \chi^2 < \chi^2_{n-1,\alpha/2}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía phải:

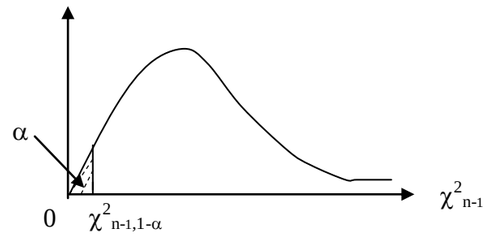
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)



* $\chi^2 \geq \chi^2_{n-1,\alpha}$ *: Bác bỏ H_o, chấp nhận H_1*

* $\chi^2 < \chi^2_{n-1,\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía trái:



* $\chi^2 \leq \chi^2_{n-1,1-\alpha}$ *: Bác bỏ H_o, chấp nhận H_1*

* $\chi^2 > \chi^2_{n-1,1-\alpha}$ *: Chưa đủ cơ sở bác bỏ H_o.*

Ví dụ: Cam kết hợp đồng với khách hàng là phương sai trọng lượng các quả dứa không lớn hơn 0,2. Để kiểm tra, người ta chọn ngẫu nhiên 100 quả. Kết quả như trong bảng trang sau. Kiểm định cam kết trên với $\alpha = 0{,}01$.

Giả thuyết:

$H_o : \sigma^2 \leq \sigma^2_o$ (với $\sigma^2_o = 0{,}2$)

$H_1 : \sigma^2 > \sigma^2_o$

Tiêu chuẩn kiểm định: $\chi^2 = \dfrac{(n-1)s^2}{\sigma_o^2}$

Trọng lượng (x_i) kg	Số quả (n_i)
0,75	3
0,85	6
0,95	8
1,05	50
1,15	13
1,25	10
1,35	7
1,45	3

Trong đó: $s^2 = \dfrac{\sum (x_i - \bar{x})^2 f_i}{n-1} = 0{,}0211$

Với: $\bar{x} = \dfrac{\sum x_i f_i}{\sum f_i} = 1{,}087$

=> $\chi^2 = \dfrac{(n-1)s^2}{\sigma_o^2} = \dfrac{(100-1)\cdot 0{,}0211}{0{,}2} = 10{,}46$

Tra bảng: $\chi^2_{n-1,\alpha} = \chi^2_{99;0,01} = 135$

$\chi^2 < \chi^2_{n-1,\alpha}$ => Chưa đủ cơ sở bác bỏ H_o (có thể tạm chấp nhận H_o với sai lầm loại II nào đó).

5.3. Kiểm định giả thuyết và ước lượng sự khác biệt tham số hai tổng thể

5.3.1. Kiểm định giả thuyết và ước lượng sự khác biệt về Số trung bình hai tổng thể, mẫu cặp

Giả thuyết về số trung bình hai tổng thể với đối thuyết hai phía:

$H_o : \mu_x - \mu_y = D_0$ (D_0 là một hằng số)

$H_1 : \mu_x - \mu_y \neq D_0$

Giả thuyết về số trung bình hai tổng thể với đối thuyết phía phải:

$H_o : \mu_x - \mu_y \leq D_0$

$H_1 : \mu_x - \mu_y > D_0$

Giả thuyết về số trung bình hai tổng thể với đối thuyết phía trái:

$H_o : \mu_x - \mu_y \geq D_0$

$H_1 : \mu_x - \mu_y < D_0$

Mẫu cặp là cặp mẫu dữ liệu có cùng một mẫu gốc ngẫu nhiên nào đó. Thường có ba trường hợp thiết lập mẫu cặp như sau.

Mẫu cặp có thể là cặp mẫu dữ liệu của hai tác nhân đang nghiên cứu nào đó tác động lên cùng một mẫu gốc ngẫu nhiên của một tổng thể gốc. Hai tác nhân đang nghiên cứu có thể là hai phương pháp sản xuất, hai phương pháp quảng cáo, hai kiểu dáng sản phẩm, hai chiếc máy, hai công nhân, hai môi trường làm việc… Trên tổng thể gốc, chọn ngẫu nhiên n đơn vị tổng thể. Cho tác nhân thứ nhất tác động lên từng đơn vị mẫu gốc rồi đo lường kết quả (mẫu thứ nhất). Sau đó, cho tác nhân thứ hai tác động lên từng đơn vị mẫu gốc và đo lường kết quả (mẫu thứ hai). Mẫu cặp này được dùng để so sánh hai tác nhân đang nghiên cứu lên cùng tổng thể gốc.

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Mẫu cặp có thể gồm một mẫu dữ liệu ngẫu nhiên (mẫu đối chứng) được lấy từ một tổng thể gốc và một mẫu dữ liệu của một tác nhân đang nghiên cứu (mẫu của tác nhân) tác động lên cùng mẫu gốc của mẫu đối chứng. Mẫu cặp này được dùng để đánh giá tác dụng của tác nhân đang nghiên cứu lên tổng thể gốc so với trường hợp không có nó (đối chứng).

Mẫu cặp cũng có thể gồm một mẫu dữ liệu ngẫu nhiên được lấy từ một tổng thể gốc ở thời gian trước (mẫu trước) và một mẫu dữ liệu ở thời gian sau (mẫu sau) trên cùng mẫu gốc của mẫu trước. Mẫu cặp này được dùng để đánh giá sự biến động của hiện tượng theo thời gian.

So với mẫu độc lập, mẫu cặp có ưu điểm là không chịu ảnh hưởng của sự sai khác mẫu gốc ban đầu. Tuy nhiên, nhược điểm là mẫu sau thường chịu ảnh hưởng của tác nhân trước vì có chung mẫu gốc. Để hạn chế nhược điểm này, người ta thường cho tác nhân sau tác động sau một khoảng thời gian nào đó mà tác động của tác nhân trước đã giảm hẳn. Dữ liệu mẫu cặp có dạng tổng quát như sau.

Mẫu gốc	1	2	3	…	n
Mẫu thứ nhất (x_i)	x_1	x_2	x_3	…	x_n
Mẫu thứ hai (y_i)	y_1	y_2	y_3	…	y_n
$d_i = x_i - y_i$	d_1	d_2	d_3	…	d_n

Việc kiểm định được thực hiện gián tiếp qua biến $d_i = x_i - y_i$. Khi đó giả thuyết trên trở thành giả thuyết về một số trung bình tổng thể như sau.

Giả thuyết hai phía:

$H_o : \mu_d = D_0$

$H_1 : \mu_d \neq D_0$

Giả thuyết phía phải:

$H_o : \mu_d \leq D_0$

$H_1 : \mu_d > D_0$

Giả thuyết phía trái:

$H_o : \mu_d \geq D_0$

$H_1 : \mu_d < D_0$

Theo định lý giới hạn trung tâm cần lấy mẫu kích thước n ≥30. Nếu muốn lấy mẫu kích thước n < 30 thì tổng thể phải có phân phối xấp xỉ chuẩn. Tương tự mục 5.2.1:

Tiêu chuẩn kiểm định: $$t = \frac{\bar{d} - D_o}{\sqrt{\frac{s_d^2}{n}}}$$

Trong đó: $\bar{d} = \frac{\sum d_i}{n}$ và : $s_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1}$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:

* $|t| \geq t_{n-1,\, \alpha/2}$: *Bác bỏ H_o, chấp nhận H_1*
* $|t| < t_{n-1,\, \alpha/2}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía phải:

* $t \geq t_{n-1,\, \alpha}$: *Bác bỏ H_o, chấp nhận H_1*
* $t < t_{n-1,\, \alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía trái:

* $t \leq -t_{n-1,\, \alpha}$: *Bác bỏ H_o, chấp nhận H_1*
* $t > -t_{n-1,\, \alpha}$ *: Chưa đủ cơ sở bác bỏ H_o.*

Ví dụ: Người ta cho rằng phương pháp sản xuất X cho năng suất không thấp hơn phương pháp sản xuất Y. Để kiểm tra, người ta chọn 15 công nhân để thử nghiệm. Kết quả như sau:

Công nhân	Năng suất X (SP/giờ)	Năng suất Y (SP/giờ)	***Công nhân***	Năng suất X (SP/giờ)	Năng suất Y (SP/giờ)
1	10	8	***9***	7	7
2	14	13	***10***	11	12
3	8	9	***11***	8	7
4	15	14	***12***	10	10
5	12	12	***13***	13	11
6	11	10	***14***	14	12
7	9	10	***15***	13	12
8	15	12			

Biết năng suất các công nhân có phân phối chuẩn. Kiểm định với α = 0,01.

Giả thuyết: $H_o : \mu_x - \mu_y \geq D_o$ (với $D_o = 0$)

$H_1 : \mu_x - \mu_y < D_o$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Ta kiểm định thông qua mẫu $d_i = x_i - y_i$:

CN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
d_i	2	1	-1	1	0	1	-1	3	0	-1	1	0	2	2	1

Tiêu chuẩn kiểm định: $$t = \frac{\bar{d} - D_o}{\sqrt{\frac{s_d^2}{n}}}$$

Trong đó: $\bar{d} = \frac{\sum d_i}{n} = 0{,}733$, $s_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1} = 1{,}496$

=> $$t = \frac{\bar{d} - D_o}{\sqrt{\frac{s_d^2}{n}}} = \frac{0{,}733 - 0}{\sqrt{\frac{1{,}496}{15}}} = 2{,}323$$

Tra bảng: $-t_{n-1,\,\alpha} = -t_{14;0{,}01} = -2{,}624$

$t > -t_{n-1,\,\alpha}$ => Chưa đủ cơ sở bác bỏ H_o (có thể tạm chấp nhận H_o với sai lầm loại II nào đó)

**** Ước lượng sự khác biệt của $\mu_x - \mu_y$:***

Suy luận tương tự mục 4.5.1 về khoảng tin cậy của μ_x, trường hợp chưa biết trước phương sai khoảng tin cậy của $\mu_d = \mu_x - \mu_y$ sẽ là:

$$\left(\bar{d} \pm t_{n-1,\alpha/2}\sqrt{\frac{s_d^2}{n}}\right)$$

Ví dụ: Với mẫu 15 công nhân ở trên, khoảng tin cậy 99% của $\mu_x - \mu_y$ là:

$$\left(\bar{d} \pm t_{n-1,\alpha/2}\sqrt{\frac{s_d^2}{n}}\right)$$

Tra bảng $t_{n-1,\,\alpha/2} = t_{14;\,0{,}005} = 2{,}977$

Thay số: $$\left(0{,}733 \pm 2{,}977\sqrt{\frac{1{,}496}{15}}\right)$$

Hay: $$(0{,}733 \pm 2{,}977.0{,}316)$$

Hay: (-0,207; 1,673)

5.3.2. Kiểm định giả thuyết và ước lượng sự khác biệt về Số trung bình hai tổng thể, mẫu độc lập

Giả thuyết về số trung bình hai tổng thể với đối thuyết hai phía:

$H_o : \mu_x - \mu_y = D_o$ (D_o là một hằng số)

$H_1 : \mu_x - \mu_y \neq D_o$

Giả thuyết về số trung bình hai tổng thể với đối thuyết phía phải:

$H_o : \mu_x - \mu_y \leq D_o$

$H_1 : \mu_x - \mu_y > D_o$

Giả thuyết về số trung bình hai tổng thể với đối thuyết phía trái:

$H_o : \mu_x - \mu_y \geq D_o$

$H_1 : \mu_x - \mu_y < D_o$

Trên tổng thể thứ nhất lấy một mẫu ngẫu nhiên kích thước n_x. Trên tổng thể thứ hai lấy một mẫu ngẫu nhiên kích thước n_y, độc lập với việc lấy mẫu của tổng thể thứ nhất. Theo định lý giới hạn trung tâm, cần lấy các mẫu ngẫu nhiên với kích thước $n_x \geq 30$ và $n_y \geq 30$. Trường hợp muốn lấy các mẫu kích thước nhỏ hơn 30 thì các tổng thể phải có phân phối xấp xỉ chuẩn.

a. Trường hợp biết phương sai của hai tổng thể:

Cả hai số trung bình mẫu đều có phân phối chuẩn:

$$\bar{x} \sim N(\mu_x, \frac{\sigma_x^2}{n_x}) \quad \text{và} \quad \bar{y} \sim N(\mu_y, \frac{\sigma_y^2}{n_y})$$

Xét biến: $d = \bar{x} - \bar{y}$

Biến d sẽ có phân phối chuẩn: $d \sim N\left[\mu_d, \sigma_d^2\right]$

Trong đó:

$$\begin{cases} \mu_d = \mu_x - \mu_y \\ \sigma_d^2 = \sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2 = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y} \end{cases}$$

Chuyển biến d về phân phối chuẩn tắc:

$$Z = \frac{d - \mu_d}{\sigma_d} = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \sim N(0,1)$$

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Nếu H_o đúng, thì $\mu_x - \mu_y = D_o$:

$$Z = \frac{(\bar{x} - \bar{y}) - D_o}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \sim N(0,1)$$

Do đó, *tiêu chuẩn kiểm định:*

$$z = \frac{(\bar{x} - \bar{y}) - D_o}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:

* $|z| \geq z_{\alpha/2}$: *Bác bỏ H_o, chấp nhận H_1*

* $|z| < z_{\alpha/2}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía phải:

* $z \geq z_{\alpha}$: *Bác bỏ H_o, chấp nhận H_1*

* $z < z_{\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía trái:

* $z \leq - z_{\alpha}$: *Bác bỏ H_o, chấp nhận H_1*

* $z > - z_{\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

Ví dụ: Người ta cho rằng mức tiêu hao NVL cho một sản phẩm của máy X là lớn hơn hoặc bằng máy Y. Để kiểm tra, người ta sản xuất thử 20 sản phẩm trên mỗi máy. Kết quả như sau:

Máy X		Máy Y	
Mức tiêu hao NVL (x_i)	Số sản phẩm (n_{xi})	Mức tiêu hao NVL (y_i)	Số sản phẩm (n_{yi})
3,0	1	2,9	2
3,1	2	3,0	2
3,2	4	3,1	3
3,3	8	3,2	9
3,4	3	3,3	3
3,5	2	3,4	1

Biết rằng mức tiêu hao NVL cả hai máy có phân phối chuẩn. Phương sai mức tiêu hao NVL máy X là 0,017 và máy Y là 0,018. Kiểm định với $\alpha = 0,01$.

Giả thuyết:

$H_o : \mu_x - \mu_y \geq D_o$ (với $D_o = 0$)

$H_1 : \mu_x - \mu_y < D_o$

Tiêu chuẩn kiểm định:

$$z = \frac{(\bar{x} - \bar{y}) - D_o}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

Trong đó: $\bar{x} = \frac{\sum x_i f_{x_i}}{\sum f_{x_i}} = 3,28$, $\bar{y} = \frac{\sum y_i f_{y_i}}{\sum f_{y_i}} = 3,16$

=> $$z = \frac{(\bar{x} - \bar{y}) - D_o}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} = \frac{(3,28 - 3,16) - 0}{\sqrt{\frac{0,017}{20} + \frac{0,018}{20}}} = 2,92$$

Tra bảng: $- z_{\alpha} = - z_{0,01} = -2,326$

$z > - z_{\alpha}$ => Chưa đủ cơ sở bác bỏ H_o (có thể tạm chấp nhận H_o với sai lầm loại II nào đó)

**** Ước lượng sự khác biệt của $\mu_x - \mu_y$:***

Suy luận tương tự mục 4.5.1 về khoảng tin cậy của μ_x, trường hợp biết trước phương sai, khoảng tin cậy của $\mu_d = \mu_x - \mu_y$ sẽ là:

$$\left((\bar{x} - \bar{y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right)$$

Ví dụ: Với hai máy trên, khoảng tin cậy với độ tin cậy 99% của của $\mu_x - \mu_y$ là:

$$\left((\bar{x} - \bar{y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right)$$

Tra bảng $z_{\alpha/2} = z_{0,005} = 2,575$

Thay số: $$\left((3,28 - 3,16) \pm 2,575 \sqrt{\frac{0,017}{20} + \frac{0,018}{20}} \right)$$

Hay: (0,012; 0,228)

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

b. Trường hợp chưa biết phương sai của hai tổng thể:

Tương tự trường hợp trên nhưng thay phương sai tổng thể bằng phương sai mẫu và sử dụng kiểm định Welch có phân phối Student:

Tiêu chuẩn kiểm định Welch:

$$t = \frac{\bar{x} - \bar{y} - D_o}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

Bậc tự do của phân phối Student này được xác định theo công thức sau:

$$df = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{1}{n_x - 1}\left(\frac{s_x^2}{n_x}\right)^2 + \frac{1}{n_y - 1}\left(\frac{s_y^2}{n_y}\right)^2}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:

* $|t| \geq t_{df,\, \alpha/2}$: *Bác bỏ H_o, chấp nhận H_1*
* $|t| < t_{df,\, \alpha/2}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía phải:

* $t \geq t_{df,\, \alpha}$: *Bác bỏ H_o, chấp nhận H_1*
* $t < t_{df,\, \alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía trái:

* $t \leq -t_{df,\, \alpha}$: *Bác bỏ H_o, chấp nhận H_1*
* $t > -t_{df,\, \alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

Ví dụ: Người ta cho rằng mức tiêu hao NVL cho một sản phẩm của máy X là lớn hơn hoặc bằng máy Y. Để kiểm tra, người ta sản xuất thử 20 sản phẩm trên mỗi máy. Kết quả như sau:

Máy X		Máy Y	
Mức tiêu hao NVL (x_i)	Số sản phẩm (n_{xi})	Mức tiêu hao NVL (y_i)	Số sản phẩm (n_{yi})
3,0	1	2,9	2
3,1	2	3,0	2
3,2	4	3,1	3
3,3	8	3,2	9
3,4	3	3,3	3
3,5	2	3,4	1

Biết rằng mức tiêu hao NVL có phân phối chuẩn. Kiểm định với $\alpha = 0{,}01$

Giả thuyết:

$H_o : \mu_x - \mu_y \geq D_o$ (với $D_o = 0$)

$H_1 : \mu_x - \mu_y < D_o$

Tiêu chuẩn kiểm định Welch:

$$t = \frac{\bar{x} - \bar{y} - D_o}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

Trong đó: $\bar{x} = \frac{\sum x_i f_{x_i}}{\sum f_{x_i}} = 3{,}28$, $\bar{y} = \frac{\sum y_i f_{y_i}}{\sum f_{y_i}} = 3{,}16$

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2 f_{x_i}}{n_x - 1} = 0{,}0164 \qquad s_y^2 = \frac{\sum (y_i - \bar{y})^2 f_{y_i}}{n_y - 1} = 0{,}0172$$

=> $$t = \frac{\bar{x} - \bar{y} - D_o}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} = \frac{3{,}28 - 3{,}16}{\sqrt{\frac{0{,}0164}{20} + \frac{0{,}0172}{20}}} = 2{,}928$$

Bậc tự do:

$$df = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{1}{n_x - 1}\left(\frac{s_x^2}{n_x}\right)^2 + \frac{1}{n_y - 1}\left(\frac{s_y^2}{n_y}\right)^2} = \frac{\left(\frac{0{,}0164}{20} + \frac{0{,}0172}{20}\right)^2}{\frac{1}{19}\left(\frac{0{,}0164}{20}\right)^2 + \frac{1}{19}\left(\frac{0{,}0172}{20}\right)^2} \approx 38$$

Tra bảng: $-t_{df,\, \alpha} = -t_{38;\, 0{,}01} = -2{,}42$

$t > -t_{df,\, \alpha}$ => Chưa đủ cơ sở bác bỏ H_o.

**** Ước lượng sự khác biệt của $\mu_x - \mu_y$:***

Tương tự, trường hợp biết trước phương sai, khoảng tin cậy của $\mu_d = \mu_x - \mu_y$ sẽ là:

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

$$\left((\bar{x}-\bar{y}) \pm t_{df;\alpha/2}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}\right)$$

Ví dụ: Với 2 máy trên, khoảng tin cậy với độ tin cậy 99% của của μ_x - μ_y:

$$\left((\bar{x}-\bar{y}) \pm t_{df;\alpha/2}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}\right)$$

Tra bảng $t_{df,\ \alpha/2} = t_{38,\ 0,005} = 2,71$

Thay số: $\left((3,28-3,16) \pm 2,71\sqrt{\frac{0,0164}{20}+\frac{0,0172}{20}}\right)$

Hay: (0,009 ; 0,231)

Chú ý:

(1) Tiêu chuẩn kiểm định Welch thường được sử dụng cho trường hợp phương sai hai tổng thể chưa biết trước và khác nhau. Tuy nhiên, chúng nên được dùng cho cả trường hợp phương sai hai tổng thể chưa biết trước nhưng bằng nhau.

(2) Thay vì kiểm định giả thuyết H_o: μ_x - μ_y = D_o, ta có thể kiểm định giả thuyết H_o: $\mu_x = \mu_y$, trên dữ liệu $y'_i = y_i + D_o$ khi xử lý dữ liệu trên SPSS.

5.3.3. Kiểm định giả thuyết và ước lượng sự khác biệt về Tỉ lệ hai tổng thể

Giả thuyết về Tỉ lệ hai tổng thể với đối thuyết hai phía:

$H_o : p_x = p_y$

$H_1 : p_x \neq p_y$

Giả thuyết về Tỉ lệ hai tổng thể với đối thuyết phía phải:

$H_o : p_x \leq p_y$

$H_1 : p_x > p_y$

Giả thuyết về Tỉ lệ hai tổng thể với đối thuyết phía trái:

$H_o : p_x \geq p_y$

$H_1 : p_x < p_y$

Trên tổng thể thứ nhất lấy một mẫu ngẫu nhiên kích thước n_x. Trên tổng thể thứ hai lấy một mẫu ngẫu nhiên kích thước n_y, độc lập với việc lấy mẫu của tổng thể thứ nhất. Kích thước hai mẫu phải đạt điều kiện $n_x p_x \geq 5$, $n_x(1-p_x) \geq 5$ và $n_y p_y \geq 5$, $n_y(1-p_y) \geq 5$.

Phân phối lấy mẫu của các Tỉ lệ mẫu có phân phối chuẩn.

$$\bar{p}_x \sim N(p_x, \frac{p_x(1-p_x)}{n_x}) \quad \text{và} \quad \bar{p}_y \sim N(p_y, \frac{p_y(1-p_y)}{n_y})$$

Lập luận tương tự mục 5.3.2.a trên biến $d = \bar{p}_x - \bar{p}_y$, ta có:

$$Z = \frac{(\bar{p}_x - \bar{p}_y) - (p_x - p_y)}{\sqrt{\frac{p_x(1-p_x)}{n_x}+\frac{p_y(1-p_y)}{n_y}}} \sim N(0,1)$$

Nếu H_o đúng, $p_x = p_y = p$: $$Z = \frac{(\bar{p}_x - \bar{p}_y)}{\sqrt{\frac{p(1-p)}{n_x}+\frac{p(1-p)}{n_y}}} \sim N(0,1)$$

Do đó, *tiêu chuẩn kiểm định:* $$z = \frac{(\bar{p}_x - \bar{p}_y)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}}$$

Trong đó: $$\bar{p} = \frac{\bar{p}_x n_x + \bar{p}_y n_y}{n_x + n_y}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:

* $|z| \geq z_{\alpha/2}$: *Bác bỏ H_o, chấp nhận H_1*

* $|z| < z_{\alpha/2}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía phải:

* $z \geq z_\alpha$: *Bác bỏ H_o, chấp nhận H_1*

* $z < z_\alpha$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía trái:

* $z \leq -z_\alpha$: *Bác bỏ H_o, chấp nhận H_1*

* $z > -z_\alpha$ *: Chưa đủ cơ sở bác bỏ Ho.*

Ví dụ: Một vị quản đốc cho rằng tỉ lệ sản phẩm hỏng của máy X là không nhỏ hơn máy Y. Để kiểm tra, người ta sản xuất 400 sản phẩm trên máy X và 500 sản phẩm trên máy Y. Kết quả cả hai máy đều có 3 sản phẩm hỏng. Kiểm định với α = 0,01.

Giả thuyết: $H_o : p_x \geq p_y$

$H_1 : p_x < p_y$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Tiêu chuẩn kiểm định: $$z = \frac{(\bar{p}_x - \bar{p}_y)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}}$$

Trong đó: $$\bar{p}_x = \frac{3}{400} = 0{,}0075$$

$$\bar{p}_y = \frac{3}{500} = 0{,}006$$

$$\bar{p} = \frac{\bar{p}_x n_x + \bar{p}_y n_y}{n_x + n_y} = \frac{0{,}0075 \cdot 400 + 0{,}006 \cdot 500}{400 + 500} = 0{,}0067$$

=> $$z = \frac{0{,}0075 - 0{,}006}{\sqrt{0{,}006(1-0{,}006)\left(\frac{1}{400}+\frac{1}{500}\right)}} = 0{,}27$$

Tra bảng: $- z_\alpha = - z_{0,01} = - 2{,}326$

$z > - z_\alpha$ => Chưa đủ cơ sở bác bỏ H_o.

**** Ước lượng sự khác biệt của p_x - p_y :***

Suy luận tương tự như mục 4.5.1 trên biến $d = \bar{p}_x - \bar{p}_y$, *khoảng tin cậy của* p_x - p_y *là:*

$$\left((\bar{p}_x - \bar{p}_y) \pm z_{\alpha/2}\sqrt{\frac{\bar{p}_x(1-\bar{p}_x)}{n_x}+\frac{\bar{p}_y(1-\bar{p}_y)}{n_y}}\right)$$

Ví dụ: Với hai máy trên, khoảng tin cậy với độ tin cậy 99% của p_x - p_y:

$$\left((\bar{p}_x - \bar{p}_y) \pm z_{\alpha/2}\sqrt{\frac{\bar{p}_x(1-\bar{p}_x)}{n_x}+\frac{\bar{p}_y(1-\bar{p}_y)}{n_y}}\right)$$

Tra bảng $z_{\alpha/2} = z_{0,005} = 2{,}575$

Thay số:

$$\left((0{,}0075 - 0{,}006) \pm 2{,}575\sqrt{\frac{0{,}0075(1-0{,}0075)}{400}+\frac{0{,}006(1-0{,}006)}{500}}\right)$$

Hay: (-1,2 ; 1,5) %

5.3.4. Kiểm định giả thuyết về Phương sai hai tổng thể

Giả thuyết về Phương sai hai tổng thể với đối thuyết hai phía:

$H_o : \sigma^2_x = \sigma^2_y$

$H_1 : \sigma^2_x \neq \sigma^2_y$

Giả thuyết về Phương sai hai tổng thể với đối thuyết phía phải:

$H_o : \sigma^2_x \leq \sigma^2_y$

$H_1 : \sigma^2_x > \sigma^2_y$

Giả thuyết về Phương sai hai tổng thể với đối thuyết phía trái:

$H_o : \sigma^2_x \geq \sigma^2_y$

$H_1 : \sigma^2_x < \sigma^2_y$

Để kiểm định, *lấy hai mẫu ngẫu nhiên độc lập kích thước* n_x, n_y. Nếu hai tổng thể có phân phối chuẩn, theo phân phối lấy mẫu của phương sai mẫu trên mỗi mẫu, ta có:

$$\frac{(n_x-1)s_x^2}{\sigma_x^2} \sim \chi^2_{n_x-1} \qquad \text{và} \qquad \frac{(n_y-1)s_y^2}{\sigma_y^2} \sim \chi^2_{n_y-1}$$

Theo phân phối Fisher-Snedecor và nếu H_o đúng, $\sigma^2_x = \sigma^2_y$:

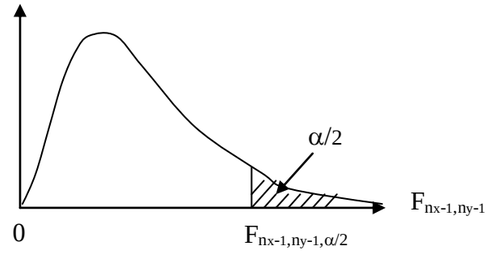
$$F_x = \frac{\frac{\chi^2_{n_x-1}}{n_x-1}}{\frac{\chi^2_{n_y-1}}{n_y-1}} = \frac{s_x^2}{s_y^2} \sim F_{n_x-1,n_y-1}$$

Tương tự: $$F_y = \frac{s_y^2}{s_x^2} \sim F_{n_y-1,n_x-1}$$

Do đó, *tiêu chuẩn kiểm định:* $$F_x = \frac{s_x^2}{s_y^2} \quad \textit{hoặc} \quad F_y = \frac{s_y^2}{s_x^2}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:



Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Nếu $s_x^2 \geq s_y^2$

* $F_x \geq F_{n_x-1,n_y-1,\alpha/2}$: *Bác bỏ H_o, chấp nhận H_1*

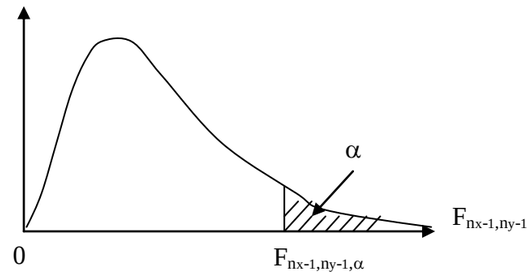
* $F_x < F_{n_x-1,n_y-1,\alpha/2}$: *Chưa đủ cơ sở bác bỏ H_o.*

Nếu $s_y^2 > s_x^2$

* $F_y \geq F_{n_y-1,n_x-1,\alpha/2}$: *Bác bỏ H_o, chấp nhận H_1*

* $F_y < F_{n_y-1,n_x-1,\alpha/2}$: *Chưa đủ cơ sở bác bỏ H_o.*

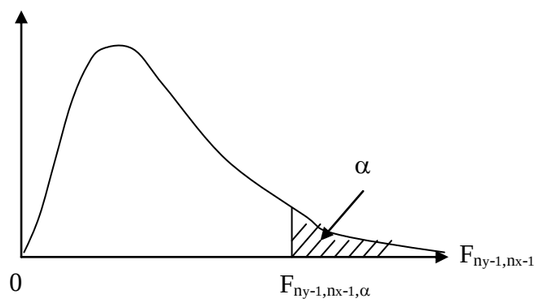
+ Kiểm định phía phải:



* $F_x \geq F_{n_x-1,n_y-1,\alpha}$: *Bác bỏ H_o, chấp nhận H_1*

* $F_x < F_{n_x-1,n_y-1,\alpha}$: *Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía trái:



* $F_y \geq F_{n_y-1,n_x-1,\alpha}$: *Bác bỏ H_o, chấp nhận H_1*

* $F_y < F_{n_y-1,n_x-1,\alpha}$: *Chưa đủ cơ sở bác bỏ Ho.*

Ví dụ: Vị quản đốc cho rằng độ lệch chuẩn trọng lượng sản phẩm của máy X là nhỏ hơn máy Y. Để kiểm tra, người ta sản xuất thử 20 sản phẩm trên mỗi máy. Kết quả như sau:

Máy X		Máy Y	
Trọng lượng SP (x_i)	Số sản phẩm (n_{xi})	Trọng lượng SP (y_i)	Số sản phẩm (n_{yi})
3,0	1	2,9	2
3,1	2	3,0	2
3,2	4	3,1	3
3,3	8	3,2	9
3,4	3	3,3	3
3,5	2	3,4	1

Biết rằng mức tiêu hao NVL cả hai máy có phân phối chuẩn. Kiểm định với $\alpha = 0,01$

Giả thuyết:

$H_o : \sigma^2_x \geq \sigma^2_y$

$H_1 : \sigma^2_x < \sigma^2_y$

Tiêu chuẩn kiểm định: $F_y = \dfrac{s_y^2}{s_x^2}$

Trong đó: $\bar{x} = \dfrac{\sum x_i f_{x_i}}{\sum f_{x_i}} = 3,28$, $\bar{y} = \dfrac{\sum y_i f_{y_i}}{\sum f_{y_i}} = 3,16$

$s_x^2 = \dfrac{\sum (x_i - \bar{x})^2 f_{x_i}}{n_x - 1} = 0,0164$ $\quad s_y^2 = \dfrac{\sum (y_i - \bar{y})^2 f_{y_i}}{n_y - 1} = 0,0172$

=> $F_y = \dfrac{s_y^2}{s_x^2} = \dfrac{0,0172}{0,0164} = 1,05$

Tra bảng: $F_{ny-1,\ nx-1,\ \alpha} = F_{19,\ 19,\ 0,01} = 2,99$

$F_y < F_{n_y-1,n_x-1,\alpha}$: Chưa đủ cơ sở bác bỏ H_o. Vậy nhận định đưa ra là không đúng.

5.4. Các kiểm định tham số khác

5.4.1. Kiểm định giả thuyết về mối liên hệ tương quan giữa hai tiêu thức định lượng (kiểm định Pearson)

Giả thuyết hai phía:

$H_o : \rho = 0$ (ρ: hệ số Pearson tổng thể)

$H_1 : \rho \neq 0$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Giả thuyết một phía thuận:

$H_o : \rho \le 0$

$H_1 : \rho > 0$

Giả thuyết một phía nghịch:

$H_o : \rho \ge 0$

$H_1 : \rho < 0$

Lấy mẫu ngẫu nhiên kích thước n, thu thập dữ liệu về hai tiêu thức x, y để phục vụ việc kiểm định.

Nếu phân phối kết hợp của hai tiêu thức là phân phối chuẩn và H_o đúng, đại lượng sau có phân phối Student nới n-2 bậc tự do.

$$t = r\sqrt{\frac{n-2}{1-r^2}} \sim T_{n-2}$$

Trong đó r là hệ số tương quan Pearson:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Do đó, *tiêu chuẩn kiểm định:*

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Giả thuyết hai phía:

** $|t| \ge t_{n-2,\alpha/2}$: Bác bỏ H_o.*

** $|t| < t_{n-2,\alpha/2}$: Chưa đủ cơ sở bác bỏ H_o*

+ Giả thuyết một phía thuận:

** $t \ge t_{n-2,\alpha}$: Bác bỏ H_o.*

** $t < t_{n-2,\alpha}$: Chưa đủ cơ sở bác bỏ H_o*

+ Giả thuyết một phía nghịch:

** $t \le - t_{n-2,\alpha}$: Bác bỏ H_o.*

** $t > - t_{n-2,\alpha}$: Chưa đủ cơ sở bác bỏ H_o*

Ví dụ: Xét mẫu 18 đơn thư khiếu nại ở trang 3. Giả sử phân phối kết hợp tuổi và giá trị sản phẩm tuân theo phân phối chuẩn. Kiểm định mối liên hệ tương quan nghịch giữa tuổi và giá trị sản phẩm của khách hàng với $\alpha = 0,05$.

Giả thuyết: $H_o : \rho \ge 0$ (ρ: hệ số Pearson tổng thể)

$H_1 : \rho < 0$

Hệ số tương quan Pearson (xem lại mục 2.5.1):

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = -0,058$$

Tiêu chuẩn kiểm định:

$$t = r\sqrt{\frac{n-2}{1-r^2}} = -0,058\sqrt{\frac{18-2}{1-(-0,058)^2}} = -0,232$$

Tra bảng: $- t_{n-2,\alpha} = - t_{16,\,0,05} = - 1,746$

$t > - t_{n-2,\alpha}$: Chưa đủ cơ sở bác bỏ H_o

Vậy, tuổi và giá trị sp không có liên hệ tương quan tuyến tính nghịch.

5.4.2. Kiểm định giả thuyết về sự bằng nhau của phương sai nhiều tổng thể (kiểm định Levene)

Giả thuyết: H_o: $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$

H_1: $\exists i \ne j$ *mà* $\sigma_i^2 \ne \sigma_j^2$ $\qquad i, j \in (\overline{1,k})$

Để kiểm định, trên tổng thể i lấy mẫu ngẫu nhiên kích thước n_i. Ký hiệu x_{ij} là quan sát thứ j của mẫu i.

Nếu giả thuyết H_o đúng, đại lượng sau có phân phối Fisher-Snedecor.

$$F = \frac{\frac{1}{k-1}\sum_{i=1}^{k} (\bar{d}_i - \bar{d})^2 n_i}{\frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i} (d_{ij} - \bar{d}_i)^2} \sim F_{(k-1,n-k)}$$

Trong đó: $n = \sum_{i=1}^{k} n_i$ *: Tổng số quan sát của các mẫu*

$d_{ij} = | x_{ij} - \bar{x}_i |$ *: Sai phân tuyệt đối quan sát j của mẫu i*

$\bar{d}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} d_{ij}$ *: Sai phân tuyệt đối trung bình của mẫu i*

$\bar{d} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} d_{ij}$ *: Sai phân tuyệt đối trung bình chung của k mẫu*

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Do đó, *tiêu chuẩn kiểm định:*

$$F = \frac{\frac{1}{k-1}\sum_{i=1}^{k}(\bar{d}_i - \bar{d})^2 n_i}{\frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(d_{ij} - \bar{d}_i)^2}$$

Nếu $F \geq F_{k-1,n-k,\alpha}$ *: Bác bỏ* H_o

Nếu $F < F_{k-1,n-k,\alpha}$ *: Chưa đủ cơ sở bác bỏ* H_o

Ví dụ: Xét mẫu 18 đơn thư khiếu nại ở trang 3, kiểm định giả thuyết cho rằng ba nhóm khách hàng với yêu cầu khiếu nại khác nhau có phương sai giá trị sản phẩm khiếu nại bằng nhau.

Giả thuyết: H_o: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$

H_1: $\exists i \neq j$ *mà* $\sigma_i^2 \neq \sigma_j^2$ $\quad i, j \in (\overline{1,3})$

Yêu cầu của KH	Gía trị sp (x_{ij})	$\bar{x}_i$	d_{ij}	$\bar{d}_i$	$(d_{ij} - \bar{d}_i)^2$
B.T (1)	2.5		3.06		1.10
B.T (1)	4.5		1.06		0.90
B.T (1)	2.8		2.76		0.56
B.T (1)	6.4	5,56	0.84	2,01	1.38
B.T (1)	6.8		1.24		0.60
B.T (1)	7.4		1.84		0.03
B.T (1)	4.4		1.16		0.72
B.T (1)	9.7		4.14		4.52
Đổi (2)	1.8		3.68		0.35
Đổi (2)	12.5		7.03		15.52
Đổi (2)	3.5	5,48	1.98	3,09	1.23
Đổi (2)	5.5		0.03		9.36
Đổi (2)	8.2		2.73		0.13
Sửa (3)	10.2		3.73		1.42
Sửa (3)	4.7		1.78		0.58
Sửa (3)	9.1	6,48	2.63	2,54	0.01
Sửa (3)	5.8		0.68		3.46
Sửa (3)	2.6		3.88		1.80
Tổng	-	-	-	-	43.67

Tiêu chuẩn kiểm định:

$$F = \frac{\frac{1}{k-1}\sum_{i=1}^{k}(\bar{d}_i - \bar{d})^2 n_i}{\frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(d_{ij} - \bar{d}_i)^2}$$

$$F = \frac{\frac{1}{3-1}\left[(2{,}01-2{,}26)^2 8 + (3{,}09-2{,}46)^2 5 + (2{,}54-2{,}46)^2 5\right]}{\frac{1}{18-3}\times 43{,}67} = 0{,}62$$

Trong đó:

$$\bar{d} = \frac{\sum \bar{d}_i \cdot n_i}{\sum n_i} = \frac{2{,}01\cdot 8 + 3{,}09\cdot 5 + 2{,}54\cdot 5}{8+5+5} = 2{,}46$$

Tra bảng: $F_{k-1,\ n-k,\ \alpha} = F_{2;15;0,05} = 3{,}68$

$F < F_{k-1,n-k,\alpha}$: Chưa đủ cơ sở bác bỏ H_o

Vậy phương sai giá trị sản phẩm của 3 nhóm khiếu nại là như nhau.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Chương 6

KIỂM ĐỊNH PHI THAM SỐ

6.1. Đặc điểm của kiểm định phi tham số

Kiểm định phi tham số là loại kiểm định được xây dựng dựa trên các loại dữ liệu gián tiếp như dấu, hạng, khoảng cách, tần số thay cho dữ liệu mẫu trực tiếp. Kiểm định phi tham số không sử dụng các tham số thống kê thông thường như số trung bình, phương sai. Vì vậy, kiểm định phi tham số có thể được sử dụng với các loại dữ liệu mà việc đo lường được thực hiện trên các thang đo không chặt chẽ, hoặc không đáp ứng các điều kiện về phân phối chặt chẽ của kiểm định tham số.

Kiểm định phi tham số được sử dụng rộng rãi vì dùng được với nhiều loại dữ liệu và không đòi hỏi các điều kiện nghiêm ngặt về phân phối tổng thể. Tuy nhiên kiểm định phi tham số thường có độ chính xác thấp hơn kiểm định tham số do đó không nên quá lạm dụng.

6.2. Kiểm định giả thuyết so sánh hai hay nhiều tổng thể

6.2.1. Kiểm định giả thuyết so sánh hai tổng thể, mẫu cặp, (kiểm định hạng và dấu Wilcoxon)

Giả thuyết về sự giống nhau của hai tổng thể với đối thuyết hai phía:

H_o: $M_{e(X)} = M_{e(Y)}$ (M_e : Trung vị)

H_1: $M_{e(X)} \neq M_{e(Y)}$

Giả thuyết về sự giống nhau của hai tổng thể với đối thuyết phía phải:

H_o: $M_{e(X)} \leq M_{e(Y)}$

H_1: $M_{e(X)} > M_{e(Y)}$

Giả thuyết về sự giống nhau của hai tổng thể với đối thuyết phía trái:

H_o: $M_{e(X)} \geq M_{e(Y)}$

H_1: $M_{e(X)} < M_{e(Y)}$

Để kiểm định, lấy mẫu cặp ngẫu nhiên kích thước m (thường m > 5). Dựa vào dữ liệu mẫu cặp ta lập bảng xếp hạng các chênh lệch X-Y như sau:

* Các chênh lệch tuyệt đối $|x_i - y_i|$ được xếp hạng chung từ nhỏ đến lớn nhưng tách riêng chênh lệch âm và dương theo hai cột, không tính các chênh lệch bằng 0.

* Các chênh lệch bằng nhau sẽ nhận hạng trung bình của chúng.

* T^+, T^- : Tổng các hạng chênh lệch +, chênh lệch -.

Đơn vị mẫu gốc	Mẫu thứ nhất (X)	Mẫu thứ hai (Y)	Chênh lệch X-Y	Hạng chênh lệch (+)	Hạng chênh lệch (-)
1	x_1	y_1			
2	x_2	y_2			
3	x_3	y_3			
…	...	…			
m	x_m	y_m			
Cộng	-	-	-	T^+	T^-

Nếu H_o đúng, các tổng hạng T^+ và T^- có phân phối Wilcoxon với n bậc tự do:

$$T^+ \sim W_n \quad \text{và} \quad T^- \sim W_n$$

Trong đó:

* W_n là phân phối Wilcoxon với n bậc tự do.
* n là số chênh lệch X-Y khác 0 được xếp hạng.

Do đó, *tiêu chuẩn kiểm định là* T^+ *hoặc* T^-.

Ta kết luận với mức ý nghĩa α *(xác suất sai lầm loại I) như sau:*

+ Kiểm định hai phía: Tiêu chuẩn kiểm định $w = \min(T^+, T^-)$

* $w \leq W_{n,\alpha/2}$ *: Bác bỏ* H_o*, chấp nhận* H_1
* $w > W_{n,\alpha/2}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía phải: Tiêu chuẩn kiểm định $w = T^-$

* $w \leq W_{n,\alpha}$ *: Bác bỏ* H_o*, chấp nhận* H_1
* $w > W_{n,\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía trái: Tiêu chuẩn kiểm định $w = T^+$

* $w \leq W_{n,\alpha}$ *: Bác bỏ* H_o*, chấp nhận* H_1
* $w > W_{n,\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

Chú ý: Nếu n ≥ 10, phân phối Wilcoxon xấp xỉ phân phối chuẩn. Có thể sử dụng tiêu chuẩn kiểm định sau để thay thế:

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

$$z = \frac{T^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:

* $|z| \geq z_{\alpha/2}$: Bác bỏ H_o, chấp nhận H_1
* $|z| < z_{\alpha/2}$: Chưa đủ cơ sở bác bỏ Ho.

+ Kiểm định phía phải:

* $z \geq z_{\alpha}$: Bác bỏ H_o, chấp nhận H_1
* $z < z_{\alpha}$: Chưa đủ cơ sở bác bỏ Ho.

+ Kiểm định phía trái:

* $z \leq - z_{\alpha}$: Bác bỏ H_o, chấp nhận H_1
* $z > - z_{\alpha}$: Chưa đủ cơ sở bác bỏ Ho.

Ví dụ: Có điểm đánh giá của 8 khách hàng được chọn ngẫu nhiên về hai kiểu dáng xe thử nghiệm trên thang điểm 10 như sau:

Khách hàng	Điểm kiểu A	Điểm kiểu B	Chênh lệch (A-B)	Hạng |A-B|	Hạng (+)	Hạng (-)
1	8	9	-1	1 (3)	-	3
2	9	8	+1	2 (3)	3	-
3	6	7	-1	3 (3)	-	3
4	5	9	-4	7	-	7
5	7	9	-2	6	-	6
6	6	6	0	-	-	-
7	8	7	+1	4 (3)	3	-
8	7	8	-1	5 (3)	-	3
Tổng	-	-	-	-	6	22

(Số trong ngoặc là hạng trung bình của những chênh lệch |A-B| bằng nhau)

Kiểm định giả thuyết cho rằng kiểu A được ưa thích bằng hoặc hơn kiểu B với $\alpha = 0,05$.

Giả thuyết: H_o: $M_{e(A)} \geq M_{e(B)}$

H_1: $M_{e(A)} < M_{e(B)}$

Tiêu chuẩn kiểm định: $w = T^+ = 6$

Tra bảng: $W_{n,\alpha} = W_{7;\,0,05} = 4$

$w > W_{n,\alpha}$ => Chưa đủ cơ sở bác bỏ Ho.

6.2.2. Kiểm định giả thuyết so sánh hai tổng thể, mẫu cặp, (kiểm định dấu)

Giả thuyết về sự giống nhau của hai tổng thể với đối thuyết hai phía:

H_o: $p^+ = 0,5$ (p^+ : xác suất tổng thể X lớn hơn tổng thể Y)

H_1: $p^+ \neq 0,5$

Giả thuyết về sự giống nhau của hai tổng thể với đối thuyết phía phải:

H_o: $p^+ \leq 0,5$

H_1: $p^+ > 0,5$

Giả thuyết về sự giống nhau của hai tổng thể với đối thuyết phía trái:

H_o: $p^+ \geq 0,5$

H_1: $p^+ < 0,5$

Để kiểm định, lấy mẫu cặp ngẫu nhiên kích thước m (m > 5). Dựa vào dữ liệu mẫu cặp ta lập bảng dấu như sau:

Đơn vị mẫu gốc	Mẫu thứ nhất (X)	Mẫu thứ hai (Y)	Dấu của chênh lệch (X-Y)
1	x_1	y_1	± hoặc 0
2	x_2	y_2	± hoặc 0
3	x_3	y_3	± hoặc 0
…	…	…	…
m	x_m	y_m	± hoặc 0

Gọi K là số dấu cộng có khả năng xuất hiện trên mẫu ngẫu nhiên gồm n dấu của cột chênh lệch (không tính chênh lệch 0).

Giả sử H_o đúng. Xác suất xuất hiện một dấu cộng là $p^+ = 0,5$. *K tuân theo phân phối nhị thức* với n phép thử (n dấu) với xác suất xuất hiện một dấu cộng $p^+ = 0,5$, không đổi. Hay K ~ B(n, 0,5).

Do đó, *tiêu chuẩn kiểm định P-value:*

Đuôi trái: $P_L = P(K=0) + P(K=1) + P(K=2) + \ldots + P(K=k)$

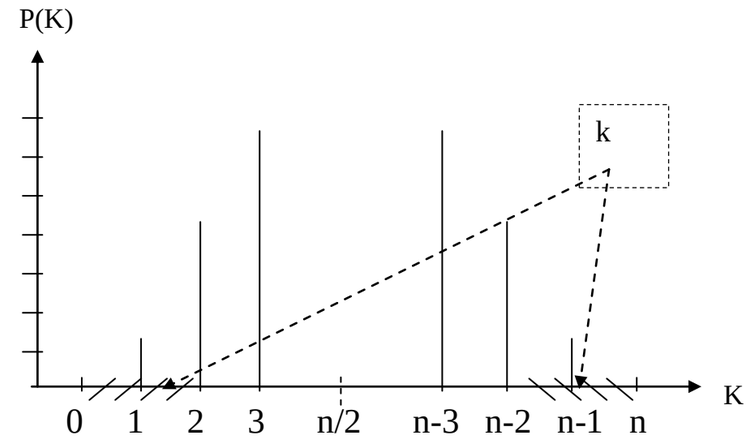
Đuôi phải: $P_R = P(K=k) + P(K=k+1) + P(K=k+2) + \ldots + P(K=n)$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau.

+ Kiểm định hai phía: Nếu k < n/2 sử dụng P_L *kiểm định đuôi trái. Nếu k≥n/2 sử dụng* P_R *kiểm định đuôi phải.*

* $2P_L \leq \alpha$ *hoặc* $2P_R \leq \alpha$ *: Bác bỏ H_o, chấp nhận H_1*
* $2P_L > \alpha$ *và* $2P_R > \alpha$ *: Chưa đủ cơ sở bác bỏ Ho.*

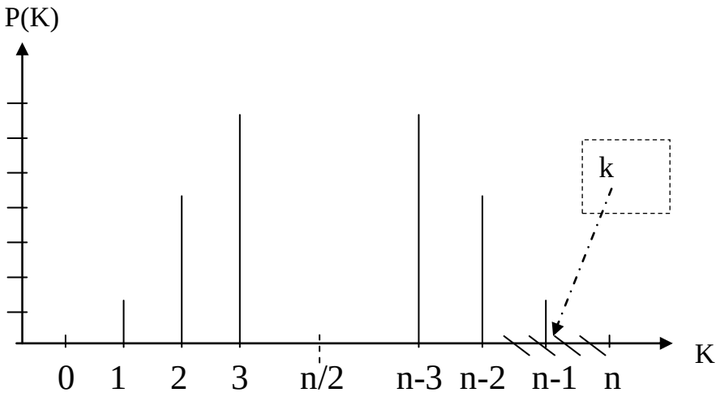
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)



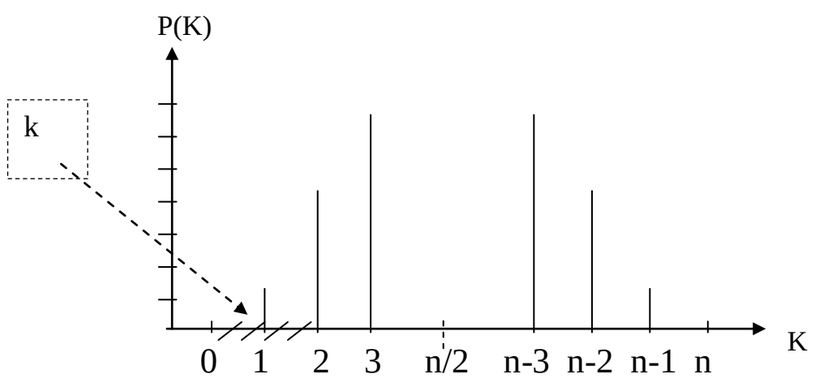
+ *Kiểm định phía phải:*

* $P_R \leq \alpha$ *: Bác bỏ* H_o*, chấp nhận* H_1

* $P_R > \alpha$ *: Chưa đủ cơ sở bác bỏ Ho.*



+ *Kiểm định phía trái:*



* $P_L \leq \alpha$ *: Bác bỏ* H_o*, chấp nhận* H_1

* $P_L > \alpha$ *: Chưa đủ cơ sở bác bỏ Ho.*

Ví dụ: Có điểm đánh giá của 8 khách hàng được chọn ngẫu nhiên về hai kiểu dáng xe thử nghiệm trên thang điểm 5 như sau:

Khách hàng	Điểm kiểu A	Điểm kiểu B	Dấu (A-B)
1	3	4	-
2	4	3	+
3	1	2	-
4	2	4	-
5	3	4	-
6	2	2	0
7	3	4	-
8	2	3	-

Kiểm định giả thuyết cho rằng kiểu A được ưa thích bằng hoặc hơn kiểu B với $\alpha = 0,05$.

Giả thuyết: H_o: $p^+ \geq 0,5$

H_1: $p^+ < 0,5$

Tổng số dấu n = 7, số dấu cộng k = 1, K ~ B(7; 0,5)

Tiêu chuẩn kiểm định:

$P_L = p(K=0) + p(K=1) = 0,0078 + 0,0547 = 0,0625$

$P_L > \alpha$ => Chưa đủ cơ sở bác bỏ Ho.. (Cụ thể, nếu bác bỏ giả thuyết này sẽ có khả năng phạm sai lầm loại I với xác suất 6,25%).

Chú ý: Nếu số dấu n > 20, phân phối nhị thức xấp xỉ phân phối chuẩn, có thể sử dụng tiêu chuẩn kiểm định z để thay thế:

$$z = \frac{\frac{k}{n} - 0,5}{\sqrt{\frac{0,5(1-0,5)}{n}}}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:

* $|z| \geq z_{\alpha/2}$: Bác bỏ H_o, chấp nhận H_1

* $|z| < z_{\alpha/2}$: Chưa đủ cơ sở bác bỏ Ho.

+ Kiểm định phía phải:

* $z \geq z_\alpha$: Bác bỏ H_o, chấp nhận H_1

* $z < z_\alpha$: Chưa đủ cơ sở bác bỏ Ho.

+ Kiểm định phía trái:

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

* $z \le -z_\alpha$: Bác bỏ H_o, chấp nhận H_1

* $z > -z_\alpha$: Chưa đủ cơ sở bác bỏ Ho.

6.2.3. Kiểm định giả thuyết so sánh hai tổng thể, mẫu độc lập

Giả thuyết về sự giống nhau của hai tổng thể với đối thuyết hai phía:

H_o: $M_{e(X)} = M_{e(Y)}$

H_1: $M_{e(X)} \neq M_{e(Y)}$

Giả thuyết về sự giống nhau của hai tổng thể với đối thuyết phía phải:

H_o: $M_{e(X)} \le M_{e(Y)}$

H_1: $M_{e(X)} > M_{e(Y)}$

Giả thuyết về sự giống nhau của hai tổng thể với đối thuyết phía trái:

H_o: $M_{e(X)} \ge M_{e(Y)}$

H_1: $M_{e(X)} < M_{e(Y)}$

Người ta thường dùng *phương pháp kiểm định hạng Mann-Whitney* như sau. Lấy các mẫu ngẫu nhiên độc lập với kích thước n_x và n_y ($n_x > 5$ và $n_y > 5$).

Lập bảng xếp hạng:

Mẫu của X	Mẫu của Y	Hạng mẫu X	Hạng mẫu Y
x_1	y_1		
x_2	y_2		
x_3	y_3		
...	...		
x_m	y_m		
-	-	T_x	T_y

* Hạng được xếp chung cho tất cả các giá trị của cả hai mẫu theo thứ tự tăng dần. Những giá trị bằng nhau sẽ nhận hạng trung bình của chúng.

* T_x, T_y : Tổng các hạng của mẫu X, mẫu Y.

Nếu H_o đúng: $$U_x = n_x n_y + \frac{n_x(n_x+1)}{2} - T_x \sim U_{n_x,n_y}$$

$$U_y = n_x n_y + \frac{n_y(n_y+1)}{2} - T_y \sim U_{n_x,n_y}$$

Trong đó, U_{n_x,n_y} là phân phối Mann-Whitney với n_x , n_y bậc tự do.

Do đó, *tiêu chuẩn kiểm định là U_x hoặc U_y.*

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía: Tiêu chuẩn kiểm định $U = \min(U_x, U_y)$

* $U \le U_{n_x,n_y,\alpha/2}$ *: Bác bỏ H_o, chấp nhận H_1*

* $U > U_{n_x,n_y,\alpha/2}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía phải: Tiêu chuẩn kiểm định U_x

* $U_x \le U_{n_x,n_y,\alpha}$ *: Bác bỏ H_o, chấp nhận H_1*

* $U_x > U_{n_x,n_y,\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía trái: Tiêu chuẩn kiểm định U_y

* $U_y \le U_{n_x,n_y,\alpha}$ *: Bác bỏ H_o, chấp nhận H_1*

* $U_y > U_{n_x,n_y,\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

Ví dụ, 10 khách hàng của Siêu thị X và 10 khách hàng của siêu thị Y được chọn ngẫu nhiên để hỏi ý kiến về mức độ hài lòng. Mỗi khách hàng được yêu cầu cho điểm về sự hài lòng trên thang 100 điểm, với 0 điểm là hoàn toàn không hài lòng và 100 điểm là hoàn toàn hài lòng. Kết quả như trong bảng sau. Kiểm định giả thuyết cho rằng sự hài lòng của khách hàng ở siêu thị X bằng hoặc nhỏ hơn ở siêu thị Y với $\alpha = 0{,}05$.

Điểm cho siêu thị X	Điểm cho siêu thị Y	Hạng mẫu X	Hạng mẫu Y
70	48	13	7
32	85	1	16
40	42	4	5
36	76	3	15
90	92	17	18
62	58	11,5	10
55	62	9	11,5
35	74	2	14
44	95	6	19
52	100	8	20
-	-	74,5	135,5

Giả thuyết: H_o: $M_{e(X)} \le M_{e(Y)}$

H_1: $M_{e(X)} > M_{e(Y)}$

Tiêu chuẩn kiểm định:

$$U_x = n_x n_y + \frac{n_x(n_x+1)}{2} - T_x = 10 \cdot 10 + \frac{10(10+1)}{2} - 74{,}5 = 80{,}5$$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Tra bảng: $U_{n_x,n_y,\alpha} = U_{10,10,0,05} = 27$

$U_x > U_{n_x,n_y,\alpha}$ => Chưa đủ cơ sở bác bỏ Ho.

Chú ý: Nếu $n_x \geq 10$ và $n_y \geq 10$, phân phối Mann-Whitney xấp xỉ phân phối chuẩn, có thể sử dụng tiêu chuẩn kiểm định sau để thay thế.

$$z = \frac{T_x - \frac{n_x(n_x+n_y+1)}{2}}{\sqrt{\frac{n_x n_y (n_x+n_y+1)}{12}}}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:

* $|z| \geq z_{\alpha/2}$: Bác bỏ H_o, chấp nhận H_1
* $|z| < z_{\alpha/2}$: Chưa đủ cơ sở bác bỏ Ho.

+ Kiểm định phía phải:

* $z \geq z_\alpha$: Bác bỏ H_o, chấp nhận H_1
* $z < z_\alpha$: Chưa đủ cơ sở bác bỏ Ho.

+ Kiểm định phía trái:

* $z \leq -z_\alpha$: Bác bỏ H_o, chấp nhận H_1
* $z > -z_\alpha$: Chưa đủ cơ sở bác bỏ Ho.

6.2.4. Kiểm định giả thuyết so sánh nhiều tổng thể

Giả thuyết:

H_o: $M_{e(1)} = M_{e(2)} = \ldots = M_{e(k)}$

H_1: $\exists\ i \neq j$ mà $M_{e(i)} \neq M_{e(j)}$ (với i, j =1, 2,…, k)

Người ta thường dùng phương pháp *kiểm định hạng Kruskal-Wallis* như sau để kiểm định loại giả thuyết trên. Trên mỗi tổng thể lấy một mẫu ngẫu nhiên độc lập với $n_i \geq 10$ (i=1, 2, … k) để kiểm định. Lập bảng xếp hạng:

Mẫu 1	Mẫu 2	…	Mẫu k	Hạng mẫu 1	Hạng mẫu 2	…	Hạng mẫu k
x_{11}	x_{21}	.	x_{k1}			.	
x_{12}	x_{22}	.	x_{k2}			.	
x_{13}	x_{23}	.	x_{k3}			.	
…	…	…	…			…	
x_{1n1}	x_{1n2}	.	$x_{k\,nk}$			.	
-	-	-		R_1	R_2	…	R_k

Trong đó:

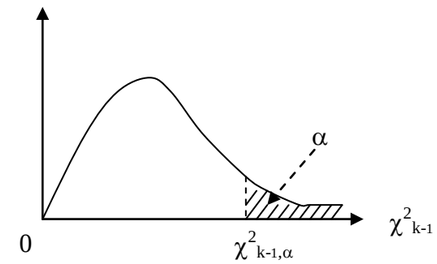
* k: Số tổng thể
* n_i : Số đơn vị tổng thể mẫu i
* $n = \sum n_i$: Tổng số đơn vị tổng thể của k mẫu
* Hạng được xếp chung cho tất cả các giá trị của cả k mẫu theo thứ tự tăng dần. Những giá trị bằng nhau sẽ nhận hạng trung bình của chúng.
* R_i : Tổng các hạng của mẫu i (i=1, 2, … k).

Nếu H_o đúng: $$H = \left[\frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}\right] - 3(n+1) \sim \chi^2_{k-1}$$

Do đó, *tiêu chuẩn kiểm định:*

$$H = \left[\frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}\right] - 3(n+1)$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):



* $H \geq \chi^2_{k-1,\,\alpha}$: *Bác bỏ H_o, chấp nhận H_1*
* $H < \chi^2_{k-1,\,\alpha}$: *Chưa đủ cơ sở bác bỏ Ho.*

Ví dụ: Một quản đốc cho rằng mức hao phí NVL (kg/SP) cho một sản phẩm của ba chiếc máy A, B, C là như nhau. Để kiểm tra, người ta sản xuất thử 7 sản phẩm trên máy A, 7 sản phẩm trên máy B và 6 sản phẩm trên máy C. Kết quả như trong bảng sau.

Kiểm định nhận định trên với α = 0,05.

Giả thuyết: H_o: $M_{e(A)} = M_{e(B)} = M_{e(C)}$

H_1: $\exists\ i \neq j$ mà $M_{e(i)} \neq M_{e(j)}$ (với i, j = A, B, C)

Tiêu chuẩn kiểm định: $$H = \left[\frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}\right] - 3(n+1)$$

$$= \frac{12}{20\cdot 21}\left(\frac{32^2}{7} + \frac{101{,}5^2}{7} + \frac{76{,}5^2}{6}\right) - 3(20+1) = 11{,}1$$

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Tra bảng: $\chi^2_{k-1,\,\alpha} = \chi^2_{2;0,05} = 5{,}99$

$H > \chi^2_{k-1,\,\alpha}$ => Bác bỏ H_o, chấp nhận H_1

Máy A		Máy B		Máy C	
Hao phí NVL/SP	Hạng	Hao phí NVL/SP	Hạng	Hao phí NVL/SP	Hạng
22,2	11	24,6	20	22,7	12
19,9	1	23,1	13	21,9	7
20,3	2,5	22,0	8	23,2	14
21,4	6	23,5	16,5	24,1	19
21,2	5	23,6	18	22,1	9,5
21,0	4	22,1	9,5	23,4	15
20,3	2,5	23,5	16,5		
Tổng hạng	32	Tổng hạng	101,5	Tổng hạng	76,5

6.3. Kiểm định giả thuyết về liên hệ giữa hai tiêu thức

6.3.1. Kiểm định giả thuyết về mối liên hệ tương quan giữa hai tiêu thức định lượng (kiểm định hạng Spearman)

Giả thuyết hai phía:

$H_o : \rho = 0$ (ρ: hệ số Pearson tổng thể)

$H_1 : \rho \neq 0$

Giả thuyết một phía thuận:

$H_o : \rho \leq 0$

$H_1 : \rho > 0$

Giả thuyết một phía nghịch:

$H_o : \rho \geq 0$

$H_1 : \rho < 0$

Trong trường hợp không biết rõ phân phối kết hợp của hai tiêu thức có tuân theo phân phối chuẩn hay không, phương pháp kiểm định tương quan hạng Spearman là thích hợp để kiểm định loại giả thuyết trên.

Lấy mẫu ngẫu nhiên kích thước n (thường n > 5) để thu thập dữ liệu theo mỗi tiêu thức. Lập bảng xếp hạng theo từng tiêu thức.

Đơn vị mẫu	Tiêu thức 1 (X)	Tiêu thức 2 (Y)	Hạng tiêu thức 1	Hạng tiêu thức 2	Chênh lệch hạng (X-Y)	d^2_i
1	x_1	y_1			d_1	d^2_1
2	x_2	y_2			d_2	d^2_2
3	x_3	y_3			d_3	d^2_3
...	...	...			...	...
n	x_n	y_n			d_n	d^2_n

Trong đó:

* Hạng được xếp riêng cho từng tiêu thức theo thứ tự tăng dần. Những giá trị bằng nhau sẽ nhận hạng trung bình của chúng.

* d_i : Chênh lệch hạng

Nếu H_o đúng: $1 - \dfrac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \sim r_n$

Trong đó, r_n là phân phối hệ số Spearman bậc n.

Do đó, *tiêu chuẩn kiểm định:* $r = 1 - \dfrac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Giả thuyết hai phía:

* $|r| \geq r_{n,\,\alpha/2}$ *: Bác bỏ* H_o.
* $|r| < r_{n,\,\alpha/2}$ *: Chưa đủ cơ sở bác bỏ* H_o

+ Giả thuyết một phía thuận:

* $r \geq r_{n,\,\alpha}$ *: Bác bỏ* H_o.
* $r < r_{n,\,\alpha}$ *: Chưa đủ cơ sở bác bỏ* H_o

+ Giả thuyết một phía nghịch:

* $r \leq -r_{n,\,\alpha}$ *: Bác bỏ* H_o.
* $r > -r_{n,\,\alpha}$ *: Chưa đủ cơ sở bác bỏ* H_o

Ví dụ: Xét mẫu 18 đơn thư khiếu nại ở trang 3. Kiểm định mối liên hệ tương quan nghịch giữa tuổi và giá trị sản phẩm của khách hàng với α = 0,05.

Giả thuyết: $H_o : \rho \geq 0$ (ρ: hệ số Spearman tổng thể)

$H_1 : \rho < 0$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Bảng xếp hạng (xem mục 2.5.1):

Đơn thư	Tuổi (x)	Gía trị SP (y)	Hạng x (R_{ix})	Hạng y (R_{iy})	R_{ix}-R_{iy} (d_i)	d_i^2
1	22	2,5	1	2	-1	1
2	26	1,8	6 (9)	1	8	64
3	25	12,5	2 (3,5)	18	-14,5	210,25
…	…	…	…	…	…	…
18	27	2,6	16 (14,5)	3	11,5	132,25
Tổng	-	-	-	-	-	1261

Tiêu chuẩn kiểm định:

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6\times 1261}{18(18^2-1} = -0{,}301$$

Tra bảng: $- r_{n,\alpha} = - r_{18,0,05} = - 0{,}399$

$r > - r_{n,\alpha}$: Chưa đủ cơ sở bác bỏ Ho..

Chú ý: Trường hợp n > 30, phân phối hệ số Spearman xấp xỉ phân phối chuẩn. Do đó, có thể chọn tiêu chuẩn kiểm định sau để thay thế.

$$z = r\sqrt{n-1}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:

* $|z| \geq z_{\alpha/2}$: Bác bỏ H_o, chấp nhận H_1
* $|z| < z_{\alpha/2}$: Chưa đủ cơ sở bác bỏ Ho.

+ Kiểm định phía phải:

* $z \geq z_{\alpha}$: Bác bỏ H_o, chấp nhận H_1
* $z < z_{\alpha}$: Chưa đủ cơ sở bác bỏ Ho.

+ Kiểm định phía trái:

* $z \leq - z_{\alpha}$: Bác bỏ H_o, chấp nhận H_1
* $z > - z_{\alpha}$: Chưa đủ cơ sở bác bỏ Ho.

6.3.2. Kiểm định giả thuyết về liên hệ giữa hai tiêu thức danh định

Giả thuyết

H_o : Hai tiêu thức là độc lập

H_1 : Hai tiêu thức có liên hệ phụ thuộc

Kiểm định Khi bình phương thường được sử dụng để kiểm định loại giả thuyết này như sau. Lấy mẫu ngẫu nhiên kích thước n. Phân tổ và lập bảng chéo theo hai tiêu thức có mặt trong giả thuyết.

Ký hiệu:

* k : Số tổ của tiêu thức 1 (X)
* m : Số tổ của tiêu thức 2 (Y)
* f_{ij} : Tần số kết hợp cả hai tiêu thức, thực tế
* f_{ix} : Tần số riêng theo tiêu thức 1 (X)
* f_{jy} : Tần số riêng theo tiêu thức 2 (Y)

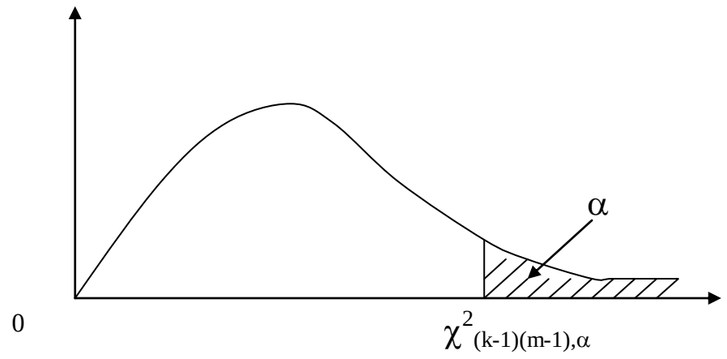
Tiêu thức 1 (X)	Tiêu thức 2 (Y)				Cộng
	y_1	y_2	…	y_m	
x_1	f_{11}	f_{12}	…	f_{1m}	f_{1x}
x_2	f_{21}	f_{21}	…	f_{2m}	f_{2x}
…	…	…	…	…	…
x_k	f_{k1}	f_{k2}	…	f_{km}	f_{kx}
Cộng	f_{1y}	f_{2y}	…	f_{my}	n

Nếu H_o đúng, tần số kết hợp cả hai tiêu thức là: $e_{ij} = \frac{f_{ix} f_{jy}}{n}$

Khi đó: $\sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \sim \chi^2_{(k-1)(m-1)}$

Do đó, *tiêu chuẩn kiểm định:* $\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):



Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

* $\chi^2 \geq \chi^2_{(k-1)(m-1),\,\alpha}$: *Bác bỏ H_o, chấp nhận H_1*

* $\chi^2 < \chi^2_{(k-1)(m-1),\,\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

***Chú ý**: Để kiểm định trên bảo đảm ý nghĩa thì tần số lý thuyết ở 80% ô trở lên phải lớn hơn hoặc bằng 5*. Do đó, kích thước mẫu cần phải khá lớn. Trường hợp không đạt được yêu cầu này, cần tăng kích thước mẫu hoặc phân tổ lại.

Ví dụ: Xét dữ liệu về mẫu 18 đơn thư khiếu nại ở trang 3. Kiểm định mối liên hệ giữa hai tiêu thức Giới tính và Yêu cầu của khách hàng.

Giả thuyết

H_o : Giới tính và Yêu cầu của khách hàng là độc lập

H_1 : Giới tính và Yêu cầu của khách hàng có liên hệ phụ thuộc

Phân tổ và lập bảng chéo theo hai tiêu thức. Tính tần số kết hợp khi 2 tiêu thức độc lập.

$$e_{ij} = \frac{f_{ix} f_{jy}}{n} \qquad (\textit{số trong ngoặc})$$

Giới tính	Yêu cầu của khách hàng			Cộng
	Bồi thường	Đổi	Sửa	
Nam	1 (4)	3 (2,5)	5 (2,5)	9
Nữ	7 (4)	2 (2,5)	0 (2,5)	9
Cộng	8	5	5	18

Tiêu chuẩn kiểm định:

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} = \frac{(1-4)^2}{4} + \frac{(3-2{,}5)^2}{2{,}5} + \ldots + \frac{(0-2{,}5)^2}{2{,}5} = 9{,}7$$

Tra bảng: $\chi^2_{(k-1)(m-1),\,\alpha} = \chi^2_{2;0,05} = 5{,}99$

$\chi^2 > \chi^2_{(k-1)(m-1),\,\alpha}$ => Bác bỏ H_o, chấp nhận H_1

Có mối liên hệ phụ thộc đáng kể giữa giới tính và loại yêu cầu của các khách hàng khi khiếu nại. Cụ thể, nữ giới có phần thiên về đòi bồi thường trong khi nam giới có phần thiên về yêu cầu sửa hoặc đổi.

6.3.3. Kiểm định giả thuyết về liên hệ giữa hai tiêu thức thứ bậc

Giả thuyết hai phía:

H_o: Hai tiêu thức không có liên hệ tương quan

H_1: Hai tiêu thức có liên hệ tương quan

Giả thuyết một phía thuận:

H_o: Hai tiêu thức không có liên hệ tương quan

H_1: Hai tiêu thức có liên hệ tương quan thuận

Giả thuyết một phía nghịch:

H_o: Hai tiêu thức không có liên hệ tương quan

H_1: Hai tiêu thức có liên hệ tương quan

a. Kiểm định Gammar:

Giả sử H_o đúng:

$$Z = \gamma\sqrt{\frac{n_c + n_d}{n(1-\gamma^2)}} \sim N(0,1)$$

Trong đó:

n: Tổng số quan sát (số đơn vị tổng thể của dữ liệu)

n_c: Số cặp quan sát có sự đồng biến trên hai tiêu thức

n_d: Số cặp quan sát có sự nghịch biến trên hai tiêu thức

$\gamma = \frac{n_c - n_d}{n_c + n_d}$: Hệ số Gammar

Do đó, *tiêu chuẩn kiểm định:* $z = \gamma\sqrt{\frac{n_c + n_d}{n(1-\gamma^2)}}$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:

* $|z| \geq z_{\alpha/2}$*: Bác bỏ H_o, chấp nhận H_1*

* $|z| < z_{\alpha/2}$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía phải:

* $z \geq z_\alpha$ *: Bác bỏ H_o, chấp nhận H_1*

* $z < z_\alpha$ *: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía trái:

* $z \leq -z_\alpha$*: Bác bỏ H_o, chấp nhận H_1*

* $z > -z_\alpha$ *: Chưa đủ cơ sở bác bỏ Ho.*

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Ví dụ: Để đánh giá mối liên hệ giữa mức độ biết đến thương hiệu và mức độ ưa thích sản phẩm của một doanh nghiệp, 200 người tiêu dùng được chọn ngẫu nhiên để thăm dò ý kiến. Kết quả được phân tổ kết hợp như sau.

Mức độ ưa thích sản phẩm	Mức độ biết đến thương hiệu		
	Kém	Khá	Tốt
Thấp	50	10	10
Trung bình	20	40	20
Cao	0	10	40

Giả thuyết: H_o: Mức độ ưa thích sp và mức độ biết đến thương hiệu không có liên hệ tương quan

H_1: Mức độ ưa thích sp và mức độ biết đến thương hiệu có liên hệ tương quan thuận

Tiêu chuẩn kiểm định:

$$z = \gamma\sqrt{\frac{n_c + n_d}{n(1-\gamma^2)}} = 0{,}77\sqrt{\frac{8700+1100}{200(1-0.77^2)}} = 8{,}45$$

Trong đó (xem mục 2.5.3): $n_c = 8700$, $n_d = 1100$, $n = 200$

$$\gamma = \frac{n_c - n_d}{n_c + n_d} = \frac{8700-1100}{8700+1100} = 0{,}77$$

Tra bảng: $Z_\alpha = Z_{0.05} = 1{,}645$

$z > Z_\alpha$: Bác bỏ H_o. Thừa nhận H_1.

Vậy có liên hệ tương quan thuận đáng kể giữa mức độ biết đến thương hiệu và mức độ ưa thích sản phẩm của doanh nghiệp nói trên.

b. Kiểm định Kendall-Tau A:

Giả sử H_o đúng:

$$z = \frac{n_c - n_d}{\sqrt{\frac{n(n-1)(2n+5)}{18}}} \sim N(0,1)$$

Do đó, *tiêu chuẩn kiểm định:*

$$z = \frac{n_c - n_d}{\sqrt{\frac{n(n-1)(2n+5)}{18}}}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I) như sau:

+ Kiểm định hai phía:

* *$|z| \geq z_{\alpha/2}$: Bác bỏ H_o, chấp nhận H_1*
* *$|z| < z_{\alpha/2}$: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía phải:

* *$z \geq z_\alpha$: Bác bỏ H_o, chấp nhận H_1*
* *$z < z_\alpha$: Chưa đủ cơ sở bác bỏ Ho.*

+ Kiểm định phía trái:

* *$z \leq -z_\alpha$: Bác bỏ H_o, chấp nhận H_1*
* *$z > -z_\alpha$: Chưa đủ cơ sở bác bỏ Ho.*

Ví dụ: Trở lại ví dụ trên.

Tiêu chuẩn kiểm định:

$$z = \frac{n_c - n_d}{\sqrt{\frac{n(n-1)(2n+5)}{18}}} = \frac{8700-1100}{\sqrt{\frac{200(200-1)(2\times200+5)}{18}}} = 8{,}03$$

Tra bảng: $Z_\alpha = Z_{0.05} = 1{,}645$

$z > Z_\alpha$: Bác bỏ H_o. Thừa nhận H_1.

Vậy có liên hệ tương quan thuận đáng kể giữa mức độ biết đến thương hiệu và mức độ ưa thích sản phẩm của doanh nghiệp nói trên.

6.4. Kiểm định giả thuyết về phân phối của một tổng thể

6.4.1. Kiểm định giả thuyết về phân phối tổng thể theo tần suất

Giả thuyết: $H_o : p_i = p_{oi}$ $\quad (i = 1, 2, \ldots k)$

$H_1 : \exists i,\ p_i \neq p_{oi}$

Trong đó:
* p_i : Tần suất (tỉ lệ) tổ i của tổng thể thực tế
* p_{oi} : Tần suất (tỉ lệ) tổ i của tổng thể theo giả thuyết
* k : Số tổ theo giả thuyết

Loại kiểm định giả thuyết này thường được gọi là kiểm định sự phù hợp (goodness of fit tests). Phương pháp kiểm định Khi bình phương thường được sử dụng. Để kiểm định, lấy mẫu ngẫu nhiên kích thước n. Lập bảng phân phối tần số thực tế và tần số lý thuyết của mẫu theo đúng k tổ của giả thuyết.

Nếu H_o đúng: $\sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} \sim \chi^2_{k-1}$

Trong đó, f_i là tần số thực tế, $e_i = n_* p_{oi}$ là tần số lý thuyết của tổ i.

Dit document is beschikbaar op studeersnel

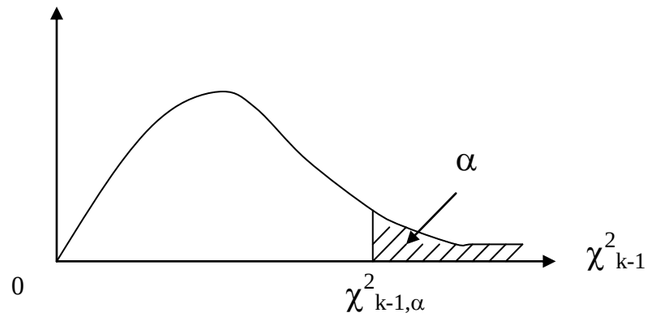
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Do đó, *tiêu chuẩn kiểm định:* $\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):

* $\chi^2 \geq \chi^2_{k-1,\,\alpha}$: *Bác bỏ H_o, chấp nhận H_1*

* $\chi^2 < \chi^2_{k-1,\,\alpha}$: *Chưa đủ cơ sở bác bỏ Ho.*



***Chú ý**: Để kiểm định trên bảo đảm ý nghĩa thì tần số lý thuyết khi H_o đúng ở 80% tổ trở lên phải lớn hơn hoặc bằng 5.*

Ví dụ: Có nhận định đưa ra là phân phối thu nhập của dân cư ở một thị trường như sau:

Mức thu nhập (triệu/người)	Tần suất % (p_{oi})
Dưới 1	10
1 - 2	20
2 - 4	40
4 - 10	15
10 - 20	10
Trên20	5

Kiểm định với $\alpha = 0{,}05$, giả thuyết:

$H_o : p_i = p_{oi}$ (i =1, 2, …6)

$H_1 : \exists i, p_i \neq p_{oi}$

Để kiểm tra, người ta điều tra mẫu ngẫu nhiên 150 người. Kết quả phân tổ, lập bảng phân phối tần số và tính tần số lý thuyết như sau:

Mức thu nhập (triệu/người)	Tần số thực tế (f_i)	Tần số khi H_o đúng ($e_i = n*p_{oi}$)
Dưới 1	25	15,0
1 - 2	38	30,0
2 - 4	52	60,0
4 - 10	18	22,5
10 - 20	12	15,0
Trên 20	5	7,5
Tổng	150	150

Tiêu chuẩn kiểm định:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

$$= \frac{(25-15{,}0)^2}{15{,}0} + \frac{(38-30{,}0)^2}{30{,}0} + \ldots + \frac{(5-7{,}5)^2}{7{,}5} = 12{,}2$$

Tra bảng: $\chi^2_{k-1,\,\alpha} = \chi^2_{5;\,0{,}05} = 11{,}07$

$\chi^2 > \chi^2_{k-1,\,\alpha}$: Bác bỏ H_o, chấp nhận H_1

6.4.2. Kiểm định giả thuyết về phân phối chuẩn của một tổng thể

Giả thuyết: H_o: Tổng thể có phân phối chuẩn

H_1: Tổng thể không có phân phối chuẩn

Phương pháp kiểm định Kolmogorov-smirnov thường được sử dụng để kiểm định loại giả thuyết này.

Xét mẫu ngẫu nhiên x_i (i=1, 2,..n) có kích thước n từ tổng thể .

Giả sử H_o đúng, x_i có phân phối chuẩn. Chuẩn hóa của x_i:

$$z_i = \frac{x_i - \bar{x}}{s_x}$$

Hàm xác suất tích lũy thực tế của z: $F_n(z) = \frac{1}{n}\sum_{i}^{n} I_{(z_i \leq z)}$

Trong đó: $I_{(zi \leq z)} = 1$ nếu $z_i \leq z$

$I_{(zi \leq z)} = 0$ nếu $z_i > z$.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Khi H_o đúng và n khá lớn thì đại lượng sau có phân phối Kolmogorov:

$$D_n = Max \mid F_n(z) - F(z) \mid \sim K$$

Trong đó, F(z) là hàm xác suất tích lũy của z có phân phối chuẩn tắc. K là phân phối Kolmogorov.

Do đó, *tiêu chuẩn kiểm định:*

$$D_n = Max \mid F_n(z) - F(z) \mid$$

Nếu $\sqrt{n}D_n \geq K_\alpha$ *: Bác bỏ* H_o

Nếu $\sqrt{n}D_n < K_\alpha$ *: Chưa đủ cơ sở bác bỏ* H_o

Một số phân vị của phân phối K.

α	0,10	0,05	0,025	0,01	0,005	0,001
K_α	1,22	1,36	1,48	1,63	1,73	1,95

Ví dụ: Xét mẫu giá trị sản phẩm trong 18 đơn thư khiếu nại ở trang 3.

Giả thuyết: H_o: Giá trị sp khiếu nại có phân phối chuẩn

H_1: Giá trị sp khiếu nại không có phân phối chuẩn

Tiêu chuẩn kiểm định:

$$D_n = Max \mid F_n(z) - F(z) \mid$$

Để tiện xác định D_n có thể dựa vào bảng tính như trang bên.

Trong đó: $\bar{x} = 6{,}02$, $s_x = 9{,}15$

$$z_i = \frac{x_i - \bar{x}}{s_x} \quad , \quad F_n(z) = \frac{1}{n}\sum_i^n I_{(z_i \leq z)}$$

=> $D_n = Max \mid F_n(z) - F(z) \mid = 0{,}268$

=> $\sqrt{n}D_n = \sqrt{18} \times 0{,}268 = 1{,}14$

Tra bảng: $K_\alpha = K_{0,05} = 1{,}36$

$\sqrt{n}D_n < K_\alpha$: Chưa đủ cơ sở bác bỏ H_o.

Vậy giá trị sản phẩm khiếu nại có phân phối chuẩn.

Bảng tính toán

Gía trị sp (x_i)	z_i	z	$F_n(z)$	F(z)	$\lvert F_n(z) - F(z) \rvert$
1.8	-0,46	-0,46	1/18	0,322	0,267
2.5	-0,38	-0,38	2/18	0,350	0,239
2.6	-0,37	-0,37	3/18	0,354	0,188
2.8	-0,35	-0,35	4/18	0,362	0,140
3.5	-0,28	-0,28	5/18	0,392	0,114
4.4	-0,18	-0,18	6/18	0,430	0,096
4.5	-0,17	-0,17	7/18	0,434	0,045
4.7	-0,14	-0,14	8/18	0,443	0,002
5.5	-0,06	-0,06	9/18	0,477	0,023
5.8	-0,02	-0,02	10/18	0,490	0,065
6.4	0,04	0,04	11/18	0,517	0,095
6.8	0,09	0,09	12/18	0,534	0,133
7.4	0,15	0,15	13/18	0,560	0,162
8.2	0,24	0,24	14/18	0,594	0,184
9.1	0,34	0,34	15/18	0,632	0,202
9.7	0,40	0,40	16/18	0,656	0,233
10.2	0,46	0,46	17/18	0,676	0,268
12.5	0,71	0,71	18/18	0,761	0,239

***Chú ý*:**

+ Để xác định $F_n(z)$ được thuận lợi cần sắp xếp z_i theo độ lớn.
+ F(z) có thể tra từ hàm NORMSDIST(z) trong EXCEL.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Chương 7

HỒI QUI

Giữa các tiêu thức định lượng của một đối tượng nghiên cứu có thể có mối liên hệ nhân quả nào đó. *Phương pháp hồi quy giúp chúng ta xác định các mô hình toán học phản ánh tốt nhất các mối liên hệ nhân quả đó.* Dựa trên các mối liên hệ và các mô hình hồi quy xác định được, ta có thể thực hiện các suy luận và dự đoán hữu ích về các hiện tượng nghiên cứu.

7.1. Hồi quy tuyến tính giữa hai tiêu thức số lượng

Dựa trên các lý thuyết kinh tế hoặc các hiểu biết thực tế về hiện tượng nghiên cứu, chọn tiêu thức gây tác động làm biến độc lập (X), tiêu thức chịu tác động làm biến phụ thuộc (Y). Nghiên cứu hồi quy giữa hai tiêu thức X, Y thường được thực hiện dựa trên mẫu dữ liệu gồm n cặp quan sát (X_i,Y_i).

Ví dụ: Để nghiên cứu mối liên hệ nhân quả giữa Quy mô gia đình (X) và Chi tiêu cho thực phẩm (Y), một mẫu 6 hộ gia đình được thu thập như sau:

X_i (người)	3	5	1	4	2	6
Y_i (triệu đồng)	0,6	1,0	0,2	1,4	0,8	1,8

7.1.1. Mô hình hồi quy tuyến tính tổng thể: *Hàm hồi quy tuyến tính tổng thể phản ảnh liên hệ nhân quả tuyến tính của một tiêu thức độc lập và một tiêu thức phụ thuộc.*

Mô hình hồi quy : $Y_i = \beta_1 + \beta_2 X_i + U_i$

Hàm hồi quy : $E(Y \mid X_i) = \beta_1 + \beta_2 X_i$

Trong đó: X : Biến (tiêu thức) độc lập

Y : Biến (tiêu thức) phụ thuộc

Y_i : Giá trị cá biệt của Y tại giá trị X_i

$E(Y|X_i)$: Giá trị trung bình (kỳ vọng) của Y tại giá trị X_i

β_1, β_2 : Các tham số của hàm hồi quy tổng thể

U_i : Yếu tố ngẫu nhiên (yếu tố nhiễu).

U_i thể hiện tác động ngẫu nhiên của tất cả các biến độc lập (biến giải thích) khác không được đưa vào mô hình đến biến phụ thuộc (biến được giải thích).

7.1.2. Mô hình hồi quy tuyến tính mẫu: *Hàm hồi quy tuyến tính mẫu là một ước lượng tốt nhất của hàm hồi quy tuyến tính tổng thể được xây dựng dựa trên một mẫu dữ liệu gồm n cặp quan sát (X_i,Y_i).*

Mô hình hồi quy mẫu: $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$

Hàm hồi qui mẫu: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

Trong đó: $\hat{Y}_i$: Ước lượng tốt nhất của $E(Y|X_i)$

$\hat{\beta}_1, \hat{\beta}_2$: Hệ số chặn, hệ số góc (ước lượng tốt nhất của β_1, β_2)

e_i : Phần dư (ước lượng tốt nhất của U_i).

7.1.3. Xác định các hệ số hàm hồi quy mẫu:

Giả sử có n cặp quan sát (X_i,Y_i), phương pháp bình phương bé nhất (Ordinary Least Square-OLS) thường được sử dụng để xác định các hệ số của hàm hồi quy tuyến tính mẫu, sao cho:

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_i)^2 => Min$$

Lấy đạo hàm riêng theo $\hat{\beta}_1, \hat{\beta}_2$ để tìm cực tiểu, ta được *công thức:*

$$\hat{\beta}_2 = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\overline{X^2} - (\overline{X})^2}$$

Và $$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}$$

Trong đó: $\overline{XY} = \dfrac{\sum_{i=1}^{n} X_i Y_i}{n}$, $\overline{X^2} = \dfrac{\sum_{i=1}^{n} X_i^2}{n}$

Ví dụ: Với dữ liệu mẫu về 6 hộ gia đình ở trên.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Hộ	X_i	Y_i	X_iY_i	X_i^2
1	3	0,6	1,8	9,0
2	5	1,0	5,0	25,0
3	1	0,2	0,2	1,0
4	4	1,4	5,6	16,0
5	2	0,8	1,6	4,0
6	6	1,8	10,8	36,0
Tổng	21	5,8	25	91
Trung bình	3,5	0,97	4,17	15,17

$$\hat{\beta}_2 = \frac{\overline{XY} - \overline{X}\cdot\overline{Y}}{\overline{X^2} - (\overline{X})^2} = \frac{4{,}17 - 3{,}5\cdot 0{,}97}{15{,}17 - 3{,}5^2} = 0{,}269$$

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2\overline{X} = 0{,}97 - 0{,}269\cdot 3{,}5 = 0{,}027$$

Hàm hồi qui mẫu: $\hat{Y}_i = 0{,}027 + 0{,}269X_i$

Ý nghĩa của các hệ số hồi quy:

- $\hat{\beta}_1$: Cho biết nếu quy mô hộ gia đình là 0 thì chi tiêu trung bình cho thực phẩm của hộ là 0,027 triệu đồng. Tuy nhiên, trị số này không có ý nghĩa thực tế.

- $\hat{\beta}_2$: Cho biết nếu quy mô hộ gia đình tăng 1 người thì chi tiêu trung bình cho thực phẩm của hộ tăng 0,269 triệu đồng.

7.1.4. Các giả thiết của mô hình hồi quy tuyến tính hai biến: Mô hình hồi quy tuyến tính được xây dựng dựa trên các giả thiết sau:

(1) Không xét đến tính ngẫu nhiên của biến độc lập trong mô hình.

(2) Kỳ vọng toán của các yếu tố nhiễu bằng 0 tại tất cả các X_i.

(3) Phương sai của các yếu tố nhiễu bằng nhau tại các X_i khác nhau.

(4) Không có tương quan giữa các yếu tố nhiễu U_i tại các X_i khác nhau.

(5) Không có tương quan giữa các yếu tố nhiễu U_i và các X_i.

(6) Yếu tố nhiễu có phân phối chuẩn: $U_i \sim N(0,\sigma^2)$

Định lý Gauss-Markov: *Với 5 giả thiết 1-5, mô hình hồi quy mẫu được xác định bằng phương pháp bình phương bé nhất là một ước lượng tốt nhất của mô hình hồi quy tổng thể (không chệch, tuyến tính, phương sai nhỏ nhất).*

Để kiểm định được các giả thuyết về mô hình hồi quy tổng thể cần có thêm giả thiết (6).

7.1.5. Hệ số xác định: $R^2 = \frac{ESS}{TSS}$

Trong đó: $TSS = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$

$$ESS = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2$$

Ý nghĩa: R^2 đo mức độ phù hợp của hàm hồi quy mẫu. Nó cho biết tỉ lệ phần trăm biến động của biến phụ thuộc được giải thích bởi hàm hồi quy mẫu.

Tính chất:

+ $0 \le R^2 \le 1$

+ $R^2 = 1$: Đường hồi quy mẫu phù hợp hoàn hảo

+ $R^2 = 0$: Giữa X và Y không có mối liên hệ

+ R^2 càng gần 1 hàm hồi quy mẫu càng phù hợp

Ví dụ: Xét dữ liệu trên, với $\hat{Y}_i = 0{,}027 + 0{,}269X_i$.

Hộ	X_i	Y_i	$\hat{Y}_i$	$(Y_i - \overline{Y})^2$	$(\hat{Y}_i - \overline{Y})^2$
1	3	0,6	0,83	0,13	0,02
2	5	1,0	1,37	0,00	0,16
3	1	0,2	0,30	0,59	0,45
4	4	1,4	1,10	0,19	0,02
5	2	0,8	0,56	0,03	0,16
6	6	1,8	1,64	0,69	0,45
Tổng cộng	21	5,8	5,8	1,63	1,26
Trung bình	3,5	0,97	0,97	0,27	0,21

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = \frac{1{,}26}{1{,}63} = 0{,}77$$

R^2 cho biết 77% biến động của chi tiêu cho thực phẩm được giải thích bởi hàm hồi quy mẫu.

7.1.6. Kiểm định sự phù hợp của mô hình hồi quy tổng thể

- Kiểm định t :

Giả thuyết: H_o: $\beta_2 = 0$

H_1: $\beta_2 \neq 0$

Tiêu chuẩn kiểm định: $t = \frac{\hat{\beta}_2}{Se(\hat{\beta}_2)}$

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Trong đó: $Se(\hat{\beta}_2) = \sqrt{\dfrac{\hat{\sigma}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$ *với:* $\hat{\sigma}^2 = \dfrac{RSS}{n-2}$

$se(\hat{\beta}_j)$: Sai số chuẩn của $\hat{\beta}_j$ (estimated standard error)

$RSS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$: Tổng bình phương phần dư

Chú ý: TSS = ESS + RSS hay RSS = TSS - ESS

Ta có thể kết luận, với sai lầm loại I mức α, như sau:

$|t| \geq t_{n-2,\alpha/2}$ *: Bác bỏ H_o, mô hình phù hợp.*

$|t| < t_{n-2,\alpha/2}$ *: Chưa đủ cơ sở bác bỏ H_o, mô hình không phù hợp.*

Ví dụ: Với hàm hồi quy $\hat{Y}_i = 0{,}027 + 0{,}269X_i$ ở trên:

Hộ	X_i	Y_i	$\hat{Y}_i$	$(X_i - \bar{X})^2$	$(Y_i - \hat{Y}_i)^2$
1	3	0,6	0,83	0,25	0,05
2	5	1,0	1,37	2,25	0,14
3	1	0,2	0,30	6,25	0,01
4	4	1,4	1,10	0,25	0,09
5	2	0,8	0,56	2,25	0,06
6	6	1,8	1,64	6,25	0,03
Tổng cộng	21	5,8	5,8	17,5	0,37
Trung bình	3,5	0,97	0,97	2,92	-

$$\hat{\sigma}^2 = \frac{RSS}{n-2} = \frac{0{,}37}{6-2} = 0{,}09$$

$$Se(\hat{\beta}_2) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}} = \sqrt{\frac{0{,}09}{17{,}5}} = 0{,}073$$

Tiêu chuẩn kiểm định: $t = \dfrac{\hat{\beta}_2}{Se(\hat{\beta}_2)} = \dfrac{0{,}269}{0{,}073} = 3{,}68$

$|t| > T_{n-2,\alpha/2} = T_{4;0,025} = 2{,}77$: Bác bỏ H_o, mô hình phù hợp.

- Kiểm định F:

Giả thuyết : $H_o: R^2 = 0$

$H_1: R^2 \neq 0$

Tiêu chuẩn kiểm định: $F = \dfrac{\dfrac{ESS}{1}}{\dfrac{RSS}{n-2}}$

Ta kết luận với sai lầm loại I mức α:

$F \geq F_{1,n-2,\alpha}$ *: Bác bỏ H_o, mô hình phù hợp.*

$F < F_{1,n-2,\alpha}$ *: Chưa đủ cơ sở bác bỏ H_o, mô hình không phù hợp.*

Ví dụ: Với hàm hồi quy $\hat{Y}_i = 0{,}027 + 0{,}269X_i$ ở trên:

$$ESS = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = 1{,}26 \quad , \quad RSS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = 0{,}37$$

$$\Rightarrow \quad F = \frac{\dfrac{ESS}{1}}{\dfrac{RSS}{n-2}} = \frac{1{,}26}{\dfrac{0{,}37}{6-2}} = 13{,}6$$

Tra bảng phân vị Fisher: $F_{1,n-2,\alpha} = F_{1;4;0,05} = 7{,}71$.

$F > F_{1,n-2,\alpha}$: Bác bỏ H_o, mô hình hồi quy phù hợp.

7.2. Hồi qui phi tuyến giữa hai tiêu thức số lượng

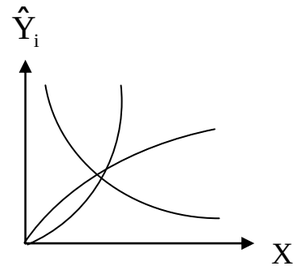
7.2.1. Mô hình hàm luỹ thừa (power)

a) Hàm hồi quy mẫu:

$$\hat{Y}_i = \hat{\beta}_1 . X_i^{\hat{\beta}_2}$$

Trong đó: $\hat{\beta}_1, \hat{\beta}_2$ các hệ số của mô hình

b) Hình dáng trên đồ thị:



c) Dạng biến đổi tuyến tính (log-log model):

$$\ln \hat{Y}_i = \ln \hat{\beta}_1 + \hat{\beta}_2 \ln X_i$$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

d) Xác định các hệ số hồi quy: Tuyến tính hóa dữ liệu bằng $X_i^* = \ln X_i$ và $Y_i^* = \ln Y_i$. Xác định hệ số $\hat{\beta}_1^*$ và $\hat{\beta}_2^*$ của mô hình tuyến tính trên dữ liệu biến đổi. Sau đó, xác định các hệ số của mô hình hàm lũy thừa ban đầu: $\hat{\beta}_1 = e^{\hat{\beta}_1^*}$ và $\hat{\beta}_2 = \hat{\beta}_2^*$.

Ví dụ: Với 6 hộ gia đình ở trên. Dữ liệu được tuyến tính hóa: $X_i^* = \ln X_i$ và $Y_i^* = \ln Y_i$.

Hộ	X_i	Y_i	X_i^*	Y_i^*	$X_i^* Y_i^*$	X_i^{*2}
1	*3*	*0,6*	1,10	-0,51	-0,56	1,21
2	*5*	*1,0*	1,61	0,00	0,00	2,59
3	*1*	*0,2*	0,00	-1,61	0,00	0,00
4	*4*	*1,4*	1,39	0,34	0,47	1,93
5	*2*	*0,8*	0,69	-0,22	-0,15	0,48
6	*6*	*1,8*	1,79	0,59	1,06	3,20
Trung bình	-	-	1,10	-0,24	0,14	1,57

$$\hat{\beta}_2^* = \frac{\overline{X^* Y^*} - \overline{X}^* \cdot \overline{Y}^*}{\overline{X^{*2}} - (\overline{X}^*)^2} = \frac{0{,}14 - 1{,}10 \cdot (-0{,}24)}{1{,}57 - 1{,}10^2} = 1{,}08$$

$$\hat{\beta}_1^* = \overline{Y}^* - \hat{\beta}_2 \overline{X}^* = (-0{,}24) - 1{,}08 \cdot 1{,}10 = -1{,}41$$

=> $\hat{\beta}_1 = e^{\hat{\beta}_1^*} = e^{-1{,}41} = 0{,}24$ và $\hat{\beta}_2 = \hat{\beta}_2^* = 1{,}08$.

Mô hình hàm lũy thừa của dữ liệu ban đầu: $\hat{Y}_i = \hat{\beta}_1 . X_i^{\hat{\beta}_2} = 0{,}24 X_i^{1{,}08}$

e) Ý nghĩa của các hệ số: Lấy đạo hàm hai vế theo X mô hình dạng biến đổi tuyến tính ta được:

$$\frac{1}{Y} \cdot \frac{dY}{dX} = \hat{\beta}_2 \frac{1}{X}$$

=>
$$\hat{\beta}_2 = \frac{\frac{dY}{Y}}{\frac{dX}{X}} = \frac{\frac{\Delta Y}{Y}}{\frac{\Delta X}{X}}$$

Như vậy: $\hat{\beta}_2$ chính là hệ số co giản của Y theo X.

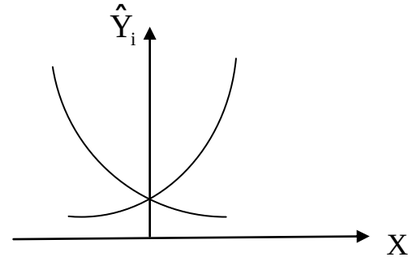
7.2.2. Mô hình hàm mũ (exponential)

a) Hàm hồi quy mẫu:

$$\hat{Y}_i = \hat{\beta}_1 . e^{\hat{\beta}_2 X_i}$$

Trong đó: $\hat{\beta}_1, \hat{\beta}_2$: Các hệ số của mô hình

b) Hình dáng trên đồ thị:



c) Dạng biến đổi tuyến tính (log-lin model):

$$\ln \hat{Y}_i = \ln \hat{\beta}_1 + \hat{\beta}_2 X_i$$

d) Xác định các hệ số mô hình:

Tuyến tính hóa dữ liệu bằng $Y_i^* = \ln Y_i$. Xác định hệ số $\hat{\beta}_1^*$ và $\hat{\beta}_2^*$ của mô hình tuyến tính trên dữ liệu biến đổi. Sau đó, xác định các hệ số của mô hình hàm mũ ban đầu: $\hat{\beta}_1 = e^{\hat{\beta}_1^*}$ và $\hat{\beta}_2 = \hat{\beta}_2^*$.

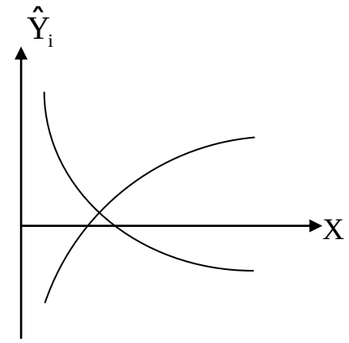
7.2.3. Mô hình hàm lôga (logarithmic model or lin-log model)

a) Hàm hồi quy mẫu:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 \ln X_i$$

Trong đó, $\hat{\beta}_1, \hat{\beta}_2$ là các hệ số của mô hình

b) Hình dáng trên đồ thị:



Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

c) Xác định các hệ số mô hình: Tuyến tính hóa dữ liệu bằng $X_i^* = \ln X_i$.

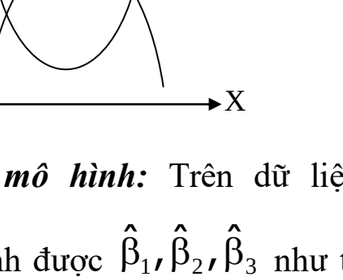
7.2.4. Mô hình hàm hypecbon (reciprocal model or inverse model)

a) Hàm hồi quy mẫu:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 \frac{1}{X_i}$$

Trong đó, $\hat{\beta}_1, \hat{\beta}_2$ là các hệ số của mô hình

b) Hình dáng trên đồ thị:



c) Xác định các hệ số mô hình: Tuyến tính hóa dữ liệu bằng $X_i^* = \ln X_i$.

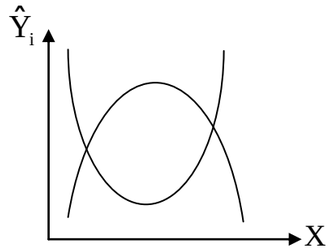
7.2.5. Mô hình hàm parabôn (quadratic model)

a) Hàm hồi quy mẫu:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 X_i^2$$

Trong đó, $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ là các hệ số của mô hình

b) Hình dáng trên đồ thị:



c) Xác định các hệ số mô hình: Trên dữ liệu tuyến tính hóa Y_i, $X_{2i} = X_i$ và $X_{3i} = X_i^2$ ta xác định được $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ như trình bày cho mô hình hồi quy tuyến tính bội.

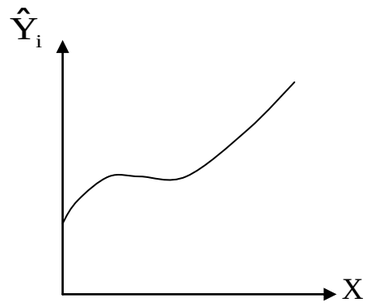
7.2.6. Mô hình hàm bậc ba (cubic model)

a) Hàm hồi quy mẫu:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 X_i^2 + \hat{\beta}_4 X_i^3$$

Trong đó, $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ là các tham số của mô hình

b) Hình dáng trên đồ thị:



c) Xác định các hệ số mô hình: Trên dữ liệu tuyến tính hóa Y_i, $X_{2i} = X_i$, $X_{3i} = X_i^2$ và $X_{4i} = X_i^3$ ta xác định được $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ như trình bày cho mô hình hồi quy tuyến tính bội.

7.2.7. Lựa chọn mô hình hồi quy: Trong trường hợp có nhiều mô hình để lựa chọn, cần dựa vào các tiêu chuẩn của một mô hình tốt sau đây.

- Mô hình càng đơn giản càng tốt.

- Mô hình giải thích được càng nhiều sự biến động của biến phụ thuộc càng tốt. Tức là R^2 (hoặc R^2 hiệu chỉnh nếu số tham số khác nhau) càng cao càng tốt.

- Mô hình vững về mặt lý thuyết. Tức là mô hình phải dựa trên một lý thuyết kinh tế nào đó và các hệ số hồi quy phải có dấu phù hợp với các lý thuyết đó.

- Mô hình phải phù hợp với thực tiễn. Tức mô hình phải có khả năng dự báo đúng các hiện tượng trong thực tế.

Ví dụ: Với dữ liệu về chi tiêu cho thực phẩm (CTCTP) và quy mô hộ gia đình (QMGD) ở trên . Kết quả hồi quy trên SPSS trong bảng sau:

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Independent: CTCTP , Dependent: QMGD							
MODEL	**Adj Rsq**	**F**	**Sigf**	**b_0**	**b_1**	**b_2**	**b_3**
LIN	,773	13,61	,021	,0267	,2686		
LOG	,740	11,39	,028	,1530	,7421		
INV	,643	7,22	,055	1,564	-1,463		
QUA	,773	5,11	,108	,0600	,2436	,0036	
CUB	,812	2,89	,268	-0,733	1,241	-0,327	,0315
POW	,839	20,82	,010	,2429	1,075		
EXP	,738	11,29	,028	,2261	,3572		

Các kết quả hồi quy thấy mô hình hàm lũy thừa (POW) và mô hình tuyến tính (LIN) là tốt hơn cả. Trong đó, hàm POW có ưu thế về R^2 hiệu chỉnh còn hàm LIN có ưu thế đơn giản. Tùy theo tính thực tiễn của mô hình mà quyết định chọn một trong hai mô hình này.

7.2. Hồi quy mối liên hệ giữa nhiều tiêu thức định lượng

Dựa trên các lý thuyết kinh tế hoặc các hiểu biết thực tế về hiện tượng nghiên cứu, chọn các tiêu thức gây tác động làm các biến độc lập (X_1, X_2, X_3,..., X_k), tiêu thức chịu tác động làm biến phụ thuộc (Y). Nghiên cứu hồi quy giữa các tiêu thức được thực hiện dựa trên mẫu dữ liệu gồm n bộ quan sát (X_{1i}, X_{2i}, X_{3i},..., X_{ki}, Y_i) .

Ví dụ: Để nghiên cứu mối liên hệ nhân quả giữa Vốn kinh doanh (VKD), Tỉ suất chi phí khuyến mãi trên doanh số (TSCPKM), Tiền lương trung bình (TLTB), Loại hình doanh nghiệp (LHDN) và Tỉ suất lợi nhuận trên vốn (TSLN), dữ liệu mẫu của 10 doanh nghiệp thương mại được thu thập như sau.

Doanh nghiệp	TSLN (%)	VKD (tỉ đồng)	TSCPKM (%)	TLTB (triệu đồng)	LHDN
1	16	2,4	0,05	2,0	Cổ phần
2	13	1,8	0,01	1,8	Tư nhân
3	12	1,5	0,03	1,9	Nhà nước
4	15	2,0	0,08	1,7	Tư nhân
5	19	2,6	0,10	2,1	Cổ phần
6	14	2,2	0,02	1,9	Tư nhân
7	17	2,5	0,04	1,7	Cổ phần
8	15	3,8	0,05	1,6	Tư nhân
9	13	1,6	0,03	1,7	Nhà nước
10	11	2,2	0,01	1,5	Nhà nước

Ta thấy Tỉ suất lợi nhuận trên vốn (Y) có khả năng phụ thuộc vào Vốn kinh doanh (X_2), Tỉ suất chi phí khuyến mãi trên doanh số (X_3), Tiền lương trung bình một nhân viên (X_4). Do đó, ta chọn Y làm biến phụ thuộc, X_2, X_3 và X_4 làm ba biến độc lập.

Mô hình hàm tuyến tính đa biến là mô hình thường được chọn nhất để nghiên cứu mối liên hệ nhân quả giữa nhiều tiêu thức số lượng vì đơn giản và khá phù hợp với nhiều hiện tượng trong thực tế.

7.3. Hồi qui giữa nhiều tiêu thức số lượng

7.3.1. Mô hình hồi quy mẫu tuyến tính đa biến:

Mô hình hồi quy mẫu: $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \ldots + \hat{\beta}_k \hat{X}_{ki} + e_i$

Hàm hồi quy mẫu: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \ldots + \hat{\beta}_k \hat{X}_{ki}$

Trong đó:

* X_{ji}: Giá trị i của biến (tiêu thức) độc lập j
* Y_i: Giá trị cá biệt của biến (tiêu thức) phụ thuộc
* $\hat{Y}_i$: Ước lượng tốt nhất của $E(Y|X_i)$
* $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \ldots \hat{\beta}_k$: Các hệ số của hàm hồi quy mẫu
* i: Quan sát thứ i
* e_i : Phần dư hay thành phần ngẫu nhiên

7.3.2. Xác định các hệ số hàm hồi quy mẫu: Tương tự mô hình hồi quy tuyến tính đơn, dựa vào phương pháp bình phương bé nhất $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \text{Min}$, lấy các đạo hàm riêng theo $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \ldots \hat{\beta}_k$ để tìm cực tiểu, ta có hệ k phương trình tuyến tính để xác định k hệ số của mô hình hồi quy mẫu.

Tuy nhiên, với trường hợp nhiều biến, công việc trên rất phức tạp. Vì vậy, một cách tổng quát, người ta thường dùng kỹ thuật ma trận để xác định các hệ số của mô hình hồi quy mẫu: $\hat{\beta} = (X'X)^{-1}X'Y$ *(xem lại lý thuyết ma trận)*

Trong đó:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \ldots \\ \hat{\beta}_k \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{21} & X_{31} & \ldots & X_{k1} \\ 1 & X_{22} & X_{32} & \ldots & X_{k2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & X_{2n} & X_{3n} & \ldots & X_{kn} \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{bmatrix}$$

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

7.3.3. Ý nghĩa các hệ số hồi quy mẫu:

- $\hat{\beta}_1$ *: Cho biết giá trị trung bình của* Y *khi không chịu tác động của các* X_j.

- $\hat{\beta}_j$ *: Cho biết mức biến động trung bình của* Y *khi* X_j *tăng 1 đơn vị còn các biến độc lập khác không đổi.*

Ví dụ: Xét dữ liệu mẫu 10 doanh nghiệp thương mại ở trên.

DN	Y_i	X_{2i}	X_{3i}	X_{4i}
1	16	2,4	0,05	2
2	13	1,8	0,01	1,8
3	12	1,5	0,03	1,9
4	15	2	0,08	1,7
5	19	2,6	0,1	2,1
6	14	2,2	0,02	1,9
7	17	2,5	0,04	1,7
8	15	3,8	0,05	1,6
9	13	1,6	0,03	1,7
10	11	2,2	0,01	1,5
Tổng	145	22,6	0,42	17,9
T.B	14,5	2,26	0,042	1,79

Hàm hồi qui bội : $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i}$

Các ma trận dữ liệu X, Y được xác định như sau:

$$X = \begin{bmatrix} 1 & 2,4 & 0,05 & 2,0 \\ 1 & 1,8 & 0,01 & 1,8 \\ 1 & 1,5 & 0,03 & 1,9 \\ 1 & 2,0 & 0,08 & 1,7 \\ 1 & 2,6 & 0,01 & 2,1 \\ 1 & 2,2 & 0,02 & 1,9 \\ 1 & 2,5 & 0,04 & 1,7 \\ 1 & 3,8 & 0,05 & 1,6 \\ 1 & 1,6 & 0,03 & 1,7 \\ 1 & 2,2 & 0,01 & 1,5 \end{bmatrix}, \quad Y = \begin{bmatrix} 16 \\ 13 \\ 12 \\ 15 \\ 19 \\ 14 \\ 17 \\ 15 \\ 13 \\ 11 \end{bmatrix}$$

Từ đó, ta tính được: $\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} 1,13 \\ 1,34 \\ 42,71 \\ 4,78 \end{bmatrix}$

Mô hình hồi quy mẫu: $\hat{Y}_i = 1,13 + 1,34X_{2i} + 42,71X_{3i} + 4,78X_{4i}$

- Ý nghĩa các hệ số hồi quy:

+ Vốn kinh doanh tăng 1 tỉ đồng thì Tỉ suất lợi nhuận tăng 1,34 điểm %.

+ Tỉ suất chi phí khuyến mãi tăng 1 điểm % thì Tỉ suất lợi nhuận tăng 42,71 điểm %.

+ Tiền lương trung bình tăng 1 triệu đồng thì Tỉ suất lợi nhuận tăng 4,78 điểm %.

7.3.4. Hệ số xác định bội:

$$R^2 = \frac{ESS}{TSS}$$

Trong đó: $TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$

$$ESS = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

Ý nghĩa: R^2 đo mức độ phù hợp của mô hình hồi quy mẫu. Nó cho biết tỉ lệ hay số phần trăm biến động của biến phụ thuộc được giải thích bởi mô hình.

** R^2 càng gần 1 mô hình càng phù hợp. R^2 càng gần 0 mô hình càng ít phù hợp.*

** R^2 chịu ảnh hưởng của số tham số trong mô hình. Do đó, người ta thường dùng thêm hệ số xác định hiệu chỉnh:*

$$\overline{R}^2 = 1 - \frac{\frac{RSS}{n-k}}{\frac{TSS}{n-1}}$$ *k: Số tham số trong mô hình*

Trong đó: $RSS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$

Ví dụ: Xét hàm $\hat{Y}_i = 1,13 + 1,34X_{2i} + 42,71X_{3i} + 4,78X_{4i}$ ở trên.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

DN	Y_i	$\hat{Y}_i$	$(Y_i-\bar{Y})^2$	$(\hat{Y}_i-\bar{Y})^2$	$(Y_i-\hat{Y}_i)^2$
1	16	16,04	2,25	2,38	0,00
2	13	12,57	2,25	3,71	0,18
3	12	13,50	6,25	0,99	2,26
4	15	15,35	0,25	0,73	0,12
5	19	18,92	20,25	19,56	0,01
6	14	14,01	0,25	0,24	0,00
7	17	14,31	6,25	0,03	7,21
8	15	16,01	0,25	2,27	1,01
9	13	12,68	2,25	3,31	0,10
10	11	11,68	12,25	7,98	0,46
Tổng	145	145,08	52,50	41,20	11,36
T.B	14,5	14,51	5,25	4,12	1,14

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2} = \frac{41,2}{52,5} = 0,78$$

$$\bar{R}^2 = 1-\frac{\frac{RSS}{n-k}}{\frac{TSS}{n-1}} = 1-\frac{\frac{11,36}{10-4}}{\frac{52,5}{10-1}} = 0,67$$

R^2 cho thấy mô hình có độ phù hợp đáng kể. Nó giải thích được khoảng 78% biến động của Tỉ suất lợi nhuận.

7.3.5. Kiểm định sự phù hợp của mô hình:

Giả thuyết: $H_o : R^2 = 0$

$H_1 : R^2 \neq 0$

Tiêu chuẩn kiểm định: $$F = \frac{\frac{ESS}{k-1}}{\frac{RSS}{n-k}}$$

Trong đó:

** k -1: Bậc tự do của ESS (số biến độc lập)*

** n – k : Bậc tự do của RSS (k số tham số của mô hình)*

Nếu: $F \geq F_{k-1,n-k,\alpha}$ *: Bác bỏ H_o, mô hình bảo đảm ý nghĩa;*

Nếu: $F < F_{k-1,n-k,\alpha}$ *: Chấp nhận H_o, mô hình không bảo đảm ý nghĩa.*

Ví dụ: Với hàm hồi quy ở trên trên,

$$F = \frac{\frac{ESS}{k-1}}{\frac{RSS}{n-k}} = \frac{\frac{41,2}{4-1}}{\frac{11,36}{10-4}} = 7,26$$

$F > F_{k-1,n-k,\alpha} = F_{3;6;0,05} = 4,76$: Bác bỏ H_o. Mô hình phù hợp.

7.3.6. Kiểm định các tham số của mô hình:

- Giả thuyết: $H_o : \beta_j = 0$

$H_1 : \beta_j \neq 0$

Tiêu chuẩn kiểm định: $$t_j = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

Trong đó: $se(\hat{\beta}) = \sqrt{\hat{\sigma}^2_{\hat{\beta}}}$

Có thể lấy $\hat{\sigma}^2_{\hat{\beta}}$ *từ đường chéo xuôi của ma trận hiệp phương sai:*

$$Cov(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} \quad \text{với:} \quad \hat{\sigma}^2 = \frac{RSS}{n-k}$$

Nếu: $|t| \geq t_{n-k,\alpha/2}$ *: Bác bỏ H_o, β_j bảo đảm ý nghĩa;*

Nếu: $|t| < t_{n-k,\alpha/2}$ *: Chấp nhận H_o, β_j không bảo đảm ý nghĩa*

Ví dụ: Với hàm hồi quy ở trên,

$$\hat{\sigma}^2 = \frac{RSS}{n-k} = \frac{11,36}{10-4} = 1,89$$

Từ ma trận hệ số X ở trên, ta tính được ma trận hiệp phương sai:

$$Cov(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} = \begin{bmatrix} 32,67 & -2,63 & 53,27 & -16,08 \\ -2,63 & 0,63 & -6,98 & 0,83 \\ 53,27 & -6,98 & 376,94 & -29,79 \\ -16,08 & 0,83 & -29,79 & 8,63 \end{bmatrix}$$

Tiêu chuẩn kiểm định β_2: $$t_2 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{1,34}{\sqrt{0,63}} = 1,68$$

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Tiêu chuẩn kiểm định β_3: $t_3 = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{42,71}{\sqrt{376,94}} = 2,2$

Tiêu chuẩn kiểm định β_4: $t_4 = \frac{\hat{\beta}_4}{se(\hat{\beta}_4)} = \frac{4,78}{\sqrt{8,63}} = 1,63$

$|t_j| < t_{n-k,\alpha/2} = t_{6;0,025} = 2,447$: Chấp nhận H_o. Cả ba tham số β_j đều không bảo đảm ý nghĩa.

7.3.7. Kiểm tra các giả thiết của mô hình hồi quy bội:

Ngoài các giả thiết tương tự mô hình hồi quy tuyến tính hai biến, mô hình hồi quy tuyến tính bội cần thêm giả thiết không có liên hệ tuyến tính giữa các biến độc lập. Hay giữa các biến độc lập không có hiện tượng đa cộng tuyến.

- Kiểm tra giả thiết không có đa cộng tuyến:

Hiện tượng đa cộng tuyến giữa các biến sẽ làm sai lệch mô hình hồi quy. Cách phát hiện hiện tượng này thường được dùng là sử dụng VIF hoặc TOL:

$$TOL_j = 1 - R_j^2$$

$$VIF_j = \frac{1}{1 - R_j^2}$$

Trong đó, R^2_j là hệ số xác định giữa biến độc lập X_j với tất cả các biến độc lập còn lại.

Nếu $VIF_j > 10$ hay $TOL_j < 0,1$ có hiện tượng đa cộng tuyến.

Những cách khắc phục hiện tượng đa cộng tuyến thông thường:

** Bỏ bớt biến có hệ số hồi quy không bảo đảm ý nghĩa.*

** Tăng thêm dữ liệu.*

** Chuyển đổi dữ liệu sang dạng ln :* $Y_{*i} = \ln Y_i$, $X_{*ji} = \ln X_{ji}$.

- Kiểm tra giả thiết phương sai đồng nhất:

Hiện tượng phương sai không đồng nhất sẽ làm sai lệch mô hình. Cách phát hiện hiện tượng này thường được dùng là sử dụng kiểm định tương quan hạng Spearman giữa $|e_i|$ với từng X_j hoặc $\hat{Y}_i$.

Nếu : $r_s > r_{n-2,\alpha/2}$ => *Có hiện tượng phương sai không đồng nhất.*

Ví dụ: Dữ liệu nói trên có kết quả hồi quy trên SPSS như sau:

	Unstandardized Coefficients		t_j	Sig.	Collinearity Statistics	
	b_j	Std. Error			Tolerance	VIF
(Constant)	1,13	5,72	,198	,85		
X_2	1,34	,80	1,68	,15	,78	1,29
X_3	42,71	19,43	2,20	,07	,65	1,55
X_4	4,78	2,94	1,63	,16	,71	1,41

Kết quả hồi quy trên cho thấy các biến X_2 và X_4 có Sig.>0,05 khá lớn, không bảo đảm ý nghĩa. Vì vậy, cần loại bớt dần các biến.

Model		Unstandardized Coefficients		t_j	Sig.	Collinearity Statistics	
		b_j	Std. Error			Tolerance	VIF
2	(Constant)	10,04	1,83	5,49	,001		
	X_2	,87	,86	1,06	***,325***	,89	1,13
	X_3	59,21	18,41	3,22	,015	,89	1,13
3	(Constant)	11,74	,88	13,32	,000		
	X_3	65,72	17,49	3,76	***,006***	1,00	1,00

Kết quả hồi quy cho thấy, trong mô hình 2 (đã loại X_4), biến X_2 có Sig.>0,05 khá lớn, không bảo đảm ý nghĩa, sẽ bị loại tiếp. Mô hình 3 chỉ còn biến X_3 có Sig.<0,05 là bảo đảm ý nghĩa.

Model Summary

Model	R	R Square
3	,80	,64

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
3	Regression	33,52	1	33,53	14,13	,006
	Residual	18,98	8	2,37		
	Total	52,50	9			

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Ta kiểm tra hiện tượng phương sai đồng nhất cho mô hình 3:

Correlations

$\lvert e_i \rvert$	$\hat{Y}_i$	X_3
Pearson Correlation	,023	,023
Sig. (2-tailed)	,950	,950

Do Sig. > 0,05, không có hiện tượng phương sai không đồng nhất.

7.4. Hồi quy với dữ liệu thuộc tính, dữ liệu thời gian

7.4.1. Hồi quy với dữ liệu định tính

Khi muốn dùng tiêu thức định tính để làm một biến độc lập giải thích cho sự biến thiên của một tiêu thức định lượng (biến phụ thuộc), trước hết ta cần lượng hóa tiêu thức thuộc tính này như sau:

Chọn một thuộc tính (biểu hiện) của tiêu thức làm thuộc tính cơ sở. Mỗi thuộc tính còn lại được lượng hóa bằng một biến giả X_i.

$$X_i = \begin{cases} 1: \text{Có thuộc tính của biến giả } X_i \\ 0: \text{Không có thuộc tính của biến giả } X_i \end{cases}$$

Ví dụ: Với dữ liệu về 10 doanh nghiệp thương mại nói trên, giả sử ta muốn nghiên cứu tác động của Tỉ suất chi phí khuyến mãi (X_1) và Loại hình doanh nghiệp (LHDN) đến Tỉ suất lợi nhuận (Y).

Ta cần lượng hóa tiêu thức Loại hình doanh nghiệp như sau:

+ Chọn loại hình doanh nghiệp tư nhân làm loại hình doanh nghiệp cơ sở.
+ Loại hình doanh nghiệp nhà nước được lượng hóa bằng biến giả X_3.
+ Loại hình doanh nghiệp cổ phần được lượng hóa bằng biến giả X_4.

$$X_3 = \begin{cases} 1: \text{Loại hình doanh nghiệp là nhà nước} \\ 0: \text{Loại hình doanh nghiệp không phải là nhà nước} \end{cases}$$

$$X_4 = \begin{cases} 1: \text{Loại hình doanh nghiệp là cổ phần} \\ 0: \text{Loại hình doanh nghiệp không phải là cổ phần} \end{cases}$$

Dữ liệu đã được lượng hóa như trong bảng sau:

Doanh nghiệp	TSLN (Y)	TSCPKM (X_2)	DNNN (X_3)	DNCP (X_4)
1	16	0,05	1	0
2	13	0,01	0	0
3	12	0,03	0	1
4	15	0,08	0	0
5	19	0,1	1	0
6	14	0,02	0	0
7	17	0,04	1	0
8	15	0,05	0	0
9	13	0,03	0	1
10	11	0,01	0	1

Ta có kết quả hồi quy như sau:

Model Summary

R	R Square	Adjusted R Square
0,97	0,95	0,92

Coefficients (Dependent Variable: Y)

	Unstandardized Coefficients		t_j	Sig.	Collinearity Statistics	
	b_j	Std. Error			Tolerance	VIF
(Constant)	12,85	0,51	25,00	0,000		
X_2	35,00	9,49	3,69	0,010	0,69	1,46
X_3	-1,67	0,55	-3,02	0,024	0,75	1,33
X_4	2,27	0,57	3,95	0,008	0,70	1,44

Cột Sig. ở bảng trên cho thấy các biến đều bảo đảm ý nghĩa. Cột VIF cho thấy không có hiện tượng đa cộng tuyến.

Correlations

$\lvert e_i \rvert$	X_2	X_3	X_4	$\hat{Y}_i$
Pearson Correlation	,19	,05	,26	,19
Sig. (2-tailed)	,60	,90	,47	,60

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Bảng trên cho thấy, kiểm định Spearman có Sig. (2-tailed) > 0,05. Không có hiện tượng phương sai không đồng nhất.

Mô hình có độ phù hợp rất cao. Trong đó, β_3 cho biết phần Tỉ suất lợi nhuận do tác dụng riêng của Loại hình doanh nghiệp nhà nước; β_4 cho biết phần Tỉ suất lợi nhuận do tác dụng riêng của Loại hình doanh nghiệp cổ phần.

7.4.2. Hồi quy với dữ liệu thời gian

Với một dãy số thời gian, các mức độ định lượng của hiện tượng nghiên cứu được chọn làm biến phụ thuộc, thời gian được dùng làm biến độc lập.

- Hàm hồi quy mẫu tổng quát:

$$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \hat{\beta}_3 X_{3t} + \ldots + \hat{\beta}_k X_{kt} + \hat{\beta}_{k+1} t$$

Trong đó, t là thứ tự thời gian.

Với mô hình trên, giá trị của biến phụ thộc ở thời gian trước thường ảnh hưởng lên giá trị của biến phụ thuộc ở các thời gian sau. Đây chính là nguyên nhân của hiện tượng tự tương quan.

- Kiểm tra hiện tượng Tự tương quan:

+ Cách phát hiện thường dùng:

Sử dụng tiêu chuẩn kiểm định Durbin-Watson:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=2}^{n} e_t^2}$$

Ta kết luận với mức ý nghĩa α cho giả thuyết hai phía:

$d < d_{L_{(n,k-1,\alpha/2)}}$ *: Có tự tương quan thuận (dương)*

$d > 4 - d_{L_{(n,k-1,\alpha/2)}}$ *: Có tự tương quan nghịch (âm)*

$d_{U_{(n,k-1,\alpha/2)}} < d < 4 - d_{U_{(n,k-1,a/2)}}$ *: Không có tự tương quan*

+ Cách khắc phục thường dùng:

** Tăng thêm dữ liệu.*

** Sử dụng hàm sai phân tổng quát bằng cách chuyển đổi dữ liệu sang:*

$$Y_{*t} = Y_t - \hat{\rho} Y_{t-1}\ ,\quad X_{*jt} = X_{jt} - \hat{\rho} X_{j(t-1)}\ ,\quad t_* = t - \hat{\rho}(t-1)$$

Trong đó: $\hat{\rho} = 1 - \frac{d}{2}$ *(j= 2,3,...k)*

Khi đó hàm hồi quy sai phân tổng quát là:

$$\hat{Y}_{*t} = \hat{\beta}_o + \hat{\beta}_2 X_{*2t} + \hat{\beta}_3 X_{*3t} + \ldots + \hat{\beta}_k X_{*kt} + \hat{\beta}_{k+1} t_*$$

Ví dụ, Có dữ liệu thu thập được trong 10 năm của một doanh nghiệp thương mại về: Tỉ suất % lợi nhuận trên vốn (TSLN), Tỉ suất % chi phí khuyến mãi trên doanh số (TSCPKM) như sau:

Năm (t)	TSLN (Y)	TSCPKM (X)
1997	14,92	0,0300
1998	16,78	0,0627
1999	16,00	0,0402
2000	16,24	0,0627
2001	16,84	0,0621
2002	17,25	0,0615
2003	17,00	0,0684
2004	17,58	0,0743
2005	17,61	0,0528
2006	18,32	0,0589

Ta thấy TSLN có khả năng phụ thuộc vào TSCPKM và Năm. Ta chọn TSLN (Y) làm biến phụ thuộc, Thời gian (t) và TSCPKM (X) làm hai biến độc lập.

Model		Unstandardized Coefficients		t_j	Sig.	Collinearity Statistics	
		b_j	Std.error			Tolerance	VIF
1	(Constant)	-456,42	100,49	-4,54	,003		
	X	21,40	11,54	1,85	*,106*	0,73	1,38
	t	,24	,05	4,68	,002	0,73	1,38
2	(Constant)	-553,27	98,06	-5,64	,000		
	t	0,29	0,05	5,81	*0,000*	1,00	1,00

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Trong mô hình 1, biến X có Sig.>0,05 nên không bảo đảm ý nghĩa, cần được loại ra khỏi mô hình. Mô hình 2, biến t có Sig.<0,05 nên bảo đảm ý nghĩa.

Model Summary

Model	R	R Square	Adjusted R Square	Durbin-Watson (d)
2	0,90	0,81	0,79	*2,75*

$d_{U_{(n,k-1,\alpha/2)}} = d_{U_{(10,1,0,05)}} = 1{,}32$, $4 - d_{U_{(n,k-1,\alpha/2)}} = 2{,}68$

$d_{L_{(n,k-1,\alpha/2)}} = d_{L_{(10,1,0,05)}} = 0{,}879$, $4 - d_{L_{(n,k-1,\alpha/2)}} = 3{,}121$

$4 - d_{U_{(n,k-1,\alpha/2)}} < d < 4 - d_{L_{(n,k-1,\alpha/2)}}$: Chưa có kết luận đáng tin cậy về hiện tượng tự tương quan.

Chương 8 PHÂN TÍCH DÃY SỐ THỜI GIAN VÀ CHỈ SỐ

Các hiện tượng kinh tế thường xuyên biến động theo thời gian do chúng chịu tác động của rất nhiều tác nhân không ngừng biến đổi xung quanh. *Các phương pháp phân tích thống kê dữ liệu thời gian sẽ giúp ta xác định được các đặc trưng biến động cơ bản của hiện tượng theo thời gian và qua đó dự đoán được các mức độ trong tương lai của hiện tượng*.

8.1. Dãy số thời gian (Time series)

8.1.1. Khái niệm, phân loại

Dãy số thời gian là một dãy các trị số (quan sát) của một hiện tượng nghiên cứu được sắp xếp theo thứ tự thời gian.

Ví dụ, dãy số thời gian về một doanh nghiệp như sau:

Tháng	1	2	3	4	5	6	7	8	9
Doanh số (tỉ đồng)	1,5	1,3	1,6	1,5	1,8	1,9	1,7	1,6	1,8

Dãy số phản ảnh mặt lượng của hiện tượng ở các thời kỳ bằng nhau được gọi là dãy số thời kỳ. Các dãy số về doanh thu, lợi nhuận, chi phí, thu nhập, sản lượng, giá trị sản xuất… của các doanh nghiệp luôn luôn là các dãy số thời kỳ.

Dãy số phản ảnh mặt lượng của hiện tượng ở các thời điểm khác nhau được gọi là dãy số thời điểm. Các dãy số về số lượng công nhân, mức tồn kho, mức dự trữ, mức vốn… của các doanh nghiệp luôn luôn là các dãy số thời điểm.

Để dãy số thời gian đáp ứng được các yêu cầu nghiên cứu biến động của hiện tượng theo thời gian cần bảo đảm tính chất so sánh được giữa các *mức độ* (trị số) của dãy số. Cụ thể:

+ Các mức độ của dãy số ở các thời gian khác nhau phải được xác định một cách thống nhất về nội dung, phương pháp tính, đơn vị tính;

+ Khoảng thời gian trong dãy số thời kỳ phải bằng nhau;

+ Khoảng cách giữa các thời gian càng gần bằng nhau càng tốt.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

8.1.2. Các thành phần của dãy số thời gian

Để tạo thuận tiện cho việc nhận thức biến động của một dãy số thời gian, người ta thường chia biến động của dãy số thời gian thành bốn thành phần sau:

- *Thành phần xu hướng:* Thành phần này thể hiện bằng biến động tăng lên hay giảm đi của hiện tượng trong một thời gian dài. Nó là kết quả tác động của những tác nhân thường xuyên, lâu dài đến hiện tượng.

- *Thành phần chu kỳ*: Thành phần này thể hiện bằng biến động tăng lên rồi giảm xuống lặp đi lặp lại theo chu kỳ trên một năm. Nó là kết quả tác động của những tác nhân có tính chu kỳ trên một năm đến hiện tượng.

- *Thành phần thời vụ:* Thành phần này thể hiện bằng biến động tăng lên rồi giảm xuống lặp đi lặp lại vào những thời gian nhất định trong năm. Nó là kết quả tác động của những tác nhân có tính thời vụ đến hiện tượng.

- *Thành phần ngẫu nhiên:* Thành phần này thể hiện bằng sự biến động tăng hay giảm bất thường, không theo qui luật nào. Nó là kết quả tác động của những tác nhân có tính ngẫu nhiên, bất thường.

Việc phân chia nói trên chỉ là tương đối về mặt nhận thức. Thành phần xu hướng và thành phần ngẫu nhiên tồn tại trong mọi dãy số thời gian. Thành phần chu kỳ và thành phần thời vụ chỉ có ở một số hiện tượng nào đó. Thành phần chu kỳ thường rất khó xác định vì tính không ổn định trong chu kỳ biến động của các hiện tượng kinh tế xã hội. Vì vậy, đôi khi thành phần này được mô tả lẫn trong thành phần xu hướng. Thành phần thời vụ là một dạng của thành phần chu kỳ nhưng được xét đến theo chu kỳ năm. Thành phần này dễ phát hiện hơn và có thể mô tả dễ dàng. Mặt khác, *cần lưu ý là, những dãy số thời gian có các trị số theo năm không có mặt thành phần thời vụ*.

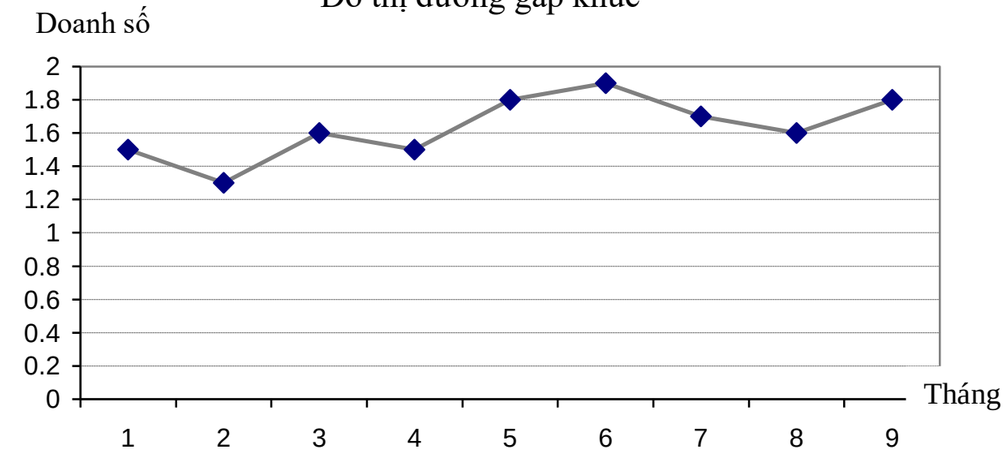
8.1.3. Mô tả dãy số thời gian bằng đồ thị (Time plots)

Đồ thị là phương pháp đầu tiên được dùng để mô tả bằng hình ảnh các đặc trưng biến động của hiện tượng qua thời gian. Phương pháp này mặc dù rất đơn giản nhưng rất hiệu quả trong việc giúp ta có được cảm nhận bằng trực giác các đặc điểm biến động cơ bản của hiện tượng qua thời gian.

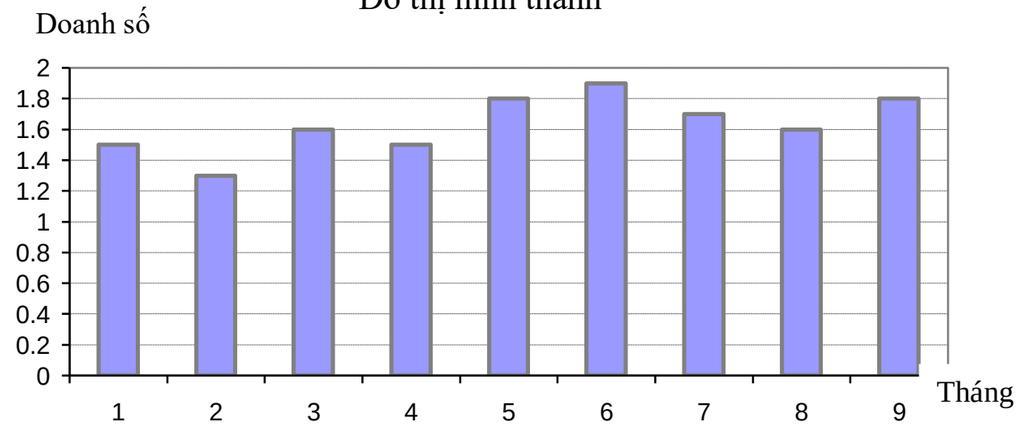
Có hai dạng đồ thị thường được dùng để mô tả dãy số thời gian là đồ thị hình thanh và đồ thị đường gấp khúc.

Ví dụ, với dãy số về doanh số bán của doanh nghiệp ở trên ta có các dạng đồ thị sau:

Đồ thị đường gấp khúc



Đồ thị hình thanh



8.2. Các chỉ tiêu mô tả dãy số thời gian

8.2.1. Số trung bình theo thời gian

Chỉ tiêu này phản ảnh trị số đại biểu của hiện tượng nghiên cứu trong suốt thời gian nghiên cứu.

a. Đối với dãy số thời kỳ:

$$\bar{y} = \frac{\sum y_i}{n}$$

Trong đó: y_i : Mức độ ở thời kỳ i, với i = 1, 2, 3, . . ., n

n : Số mức độ của dãy số

b. Đối với dãy số thời điểm:

- Trường hợp tổng quát:

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} t_1 + \frac{y_2 + y_3}{2} t_2 + \ldots + \frac{y_{n-2} + y_{n-1}}{2} t_{n-2} + \frac{y_{n-1} + y_n}{2} t_{n-1}}{t_1 + t_2 + \ldots + t_{n-2} + t_{n-1}}$$

Trong đó: y_i : Mức độ ở thời điểm i, với i = 1, 2, 3, . . ., n

t_i : Trọng số (khoảng thời gian từ thời điểm i đến thời điểm i+1)

Ví dụ, có mức tồn kho của một doanh nghiệp như sau:

Ngày	1/1	15/1	22/1	25/1	31/1
Tồn kho (tỉ đồng)	1,5	1,1	0,6	1,0	0,4

Mức tồn kho trung bình trong tháng 1:

$$\bar{y} = \frac{\frac{1,5+1,1}{2}14 + \frac{1,1+0,6}{2}7 + \frac{0,6+1,0}{2}3 + \frac{1,0+0,4}{2}6}{14+7+3+6} = 1,025 \text{ (tỉ đồng)}$$

- Trường hợp khoảng cách thời gian bằng nhau:

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + y_3 + \ldots + y_{n-1} + \frac{1}{2}y_n}{n-1}$$

Trong đó: y_i : Mức độ ở thời điểm i, với i = 1, 2, 3, . . ., n

n : Số mức độ của dãy số

Ví dụ: Có mức dự trữ hàng hoá trong các tháng đầu năm của một kho hàng như sau:

Ngày	1/1	1/ 2	1/3	1/ 4	1/5	1/6	1/7
Mức dự trữ (tỉ đồng)	1,5	1,1	0,6	1,0	0,4	1,8	1,2

Mức tồn kho trung bình trong sáu tháng đầu năm (từ 1/1 đến 30/6):

$$\bar{y} = \frac{\frac{1}{2}1,5 + 1,1 + 0,6 + 1,0 + 0,4 + 1,8 + \frac{1}{2}1,2}{7-1} = 6,25 \text{ (tỉ đồng)}$$

- Trường hợp khoảng cách thời gian không bằng nhau nhưng các mức độ không đổi trong từng khoảng thời gian:

$$\bar{y} = \frac{\sum y_i t_i}{\sum t_i}$$

Trong đó: y_i : Mức độ ở khoảng thời gian i

t_i : Khoảng thời gian tồn tại mức độ y_i

Ví dụ: Ngày 1/1 tồn kho 1,3 tỉ đồng, ngày 15/1 xuất kho 0,45 tỉ, ngày 2/2 xuất kho 0,05 tỉ, ngày 25/2 xuất kho 0,1 tỉ, ngày 10/3 nhập kho 0,4 tỉ, ngày 3/4 nhập kho 0,2 tỉ.

Từ dữ liệu trên, ta lập bảng tồn kho như sau:

Khoảng thời gian	Mức tồn kho (y_i)	Số ngày (t_i)
Từ 1/1 đến 15/1	1,30	14
Từ 15/1 đến 2/2	0,85	18
Từ 2/2 đến 25/2	0,80	23
Từ 25/2 đến 10/3	0,70	13
Từ 10/3 đến 31/3	1,10	21

(Ở đây, các thời điểm giữa ngày được chọn để tính toán. Nếu chọn đầu ngày hay cuối ngày thì kết quả có khác đôi chút)

Mức tồn kho trung bình trong quí I (từ 1/1 đến 31/3):

$$\bar{y} = \frac{\sum y_i t_i}{\sum t_i} = \frac{1,3 * 14 + 0,85 * 18 + \ldots + 1,1 * 21}{14 + 18 + \ldots + 21} = 0,94 \text{ (tỉ đồng)}$$

Chú ý: Dãy số tồn kho trong bảng trên không phải là dãy số thời kỳ vì các khoảng thời gian không bằng nhau. Rất khó thực hiện các phân tích thống kê khác trên dãy số này.

8.2.2. Lượng tăng (giảm) tuyệt đối

Chỉ tiêu này phản ảnh sự biến động tuyệt đối của hiện tượng qua thời gian. Có ba loại lượng tăng (giảm) tuyệt đối:

- *Lượng tăng (giảm) tuyệt đối liên hoàn*:

$$\partial_i = y_i - y_{i-1} \qquad \text{Với } i = 2, 3, \ldots, n$$

- *Lượng tăng (giảm) tuyệt đối định gốc*:

$$\Delta_i = y_i - y_1 \qquad \text{Với } i = 2, 3, \ldots, n$$

Ta có quan hệ: $\Delta_i = \sum_{j=2}^{i} \partial_j$ Với i = 2, 3,…,n

- *Lượng tăng (giảm) tuyệt đối trung bình*: $\bar{\partial} = \frac{\sum_{2}^{n} \partial_i}{n-1} = \frac{\Delta_n}{n-1}$

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Ví dụ, có dãy số thời gian về một doanh nghiệp:

Năm	09	10	11	12	13	14
Lợi nhuận (tỉ đ)	2	3	5	4	6	7

$$\partial_{10} = y_{11} - y_{10} = 3 - 2 = 1 \quad , \quad \partial_{13} = y_{13} - y_{12} = 6 - 4 = 2$$

$$\Delta_{12} = y_{12} - y_{09} = 4 - 2 = 2 \quad , \quad \Delta_{14} = y_{14} - y_{09} = 7 - 2 = 5$$

$$\bar{\partial} = \frac{\Delta_n}{n-1} = \frac{y_{14} - y_{09}}{6-1} = \frac{7-2}{5} = 1$$

Chú ý: Đối với dãy số thời gian có đơn vị tính %, để tránh nhầm lẫn Lượng tăng (giảm) tuyệt đối này với Tốc độ tăng (giảm) tương đối dưới đây, *người ta thường đọc chỉ tiêu này với đơn vị tính là điểm % tăng (giảm).* Ví dụ GDP năm trước tăng 7,2% năm nay tăng 8,1% , lượng tăng tuyệt đối giữa hai năm là 0,9 điểm %.

8.2.3. Tốc độ phát triển

Chỉ tiêu này phản ảnh sự biến động tương đối của hiện tượng qua thời gian. Có ba loại tốc độ phát triển sau:

- *Tốc độ phát triển liên hoàn*:

$$t_i = \frac{y_i}{y_{i-1}} \qquad \text{Với i = 2, 3,…,n}$$

- *Tốc độ phát triển định gốc*:

$$T_i = \frac{y_i}{y_1} \qquad \text{Với i = 2, 3,…,n}$$

Ta có quan hệ: $T_i = \prod_{j=2}^{i} t_j$ Với i = 2, 3,…,n

- *Tốc độ phát triển trung bình*: $\bar{t} = \sqrt[n-1]{\prod_{2}^{n} t_i} = \sqrt[n-1]{T_n}$

Ví dụ, có dãy số thời gian về một doanh nghiệp:

Năm	09	10	11	12	13	14
Lợi nhuận (tỉ đ)	2	3	5	4	6	7

$$t_{11} = \frac{y_{11}}{y_{10}} = \frac{5}{3} = 1{,}67 \quad , \quad t_{13} = \frac{y_{13}}{y_{12}} = \frac{6}{4} = 1{,}5$$

$$T_{11} = \frac{y_{11}}{y_{09}} = \frac{5}{2} = 2{,}5 \quad , \quad T_{13} = \frac{y_{13}}{y_{09}} = \frac{6}{2} = 3$$

$$\bar{t} = \sqrt[n-1]{T_n} = \sqrt[6-1]{\frac{y_{14}}{y_{09}}} = \sqrt[5]{\frac{7}{2}} = 1{,}28$$

Chú ý: Khi có dữ liệu về tốc độ phát triển trung bình từng khoảng thời gian, *tốc độ phát triển trung bình một đơn vị thời gian chung được tính theo công thức trung bình nhân có trọng số:*

$$\bar{t} = \sqrt[f_1+f_2+\ldots+f_n]{(\bar{t}_1)^{f_1}(\bar{t}_2)^{f_2} \cdots (\bar{t}_n)^{f_n}}$$

f_i *: Số đơn vị thời gian của khoảng thời gian i (quyền số)*

$\bar{t}_i$ *: Tốc độ phát triển trung bình của khoảng thời gian i*

Ví dụ, có dữ liệu về tốc độ phát triển sản lượng của một doanh nghiệp:

Khoảng thời gian	95-99	99-05	05-07	07-10
Số năm	4	6	2	3
Tốc độ phát triển trung bình một năm	1,12	1,04	1,16	1,08

Tốc độ phát triển doanh số trung bình một năm giai đoạn 1995-2010:

$$\bar{t} = \sqrt[f_1+f_2+\ldots+f_n]{(\bar{t}_1)^{f_1}(\bar{t}_2)^{f_2} \cdots (\bar{t}_n)^{f_n}}$$

$$= \sqrt[4+6+2+3]{1{,}12^4 \cdot 1{,}04^6 \cdot 1{,}16^2 \cdot 1{,}08^3} = 1{,}084$$

8.2.4. Tốc độ tăng (giảm)

Chỉ tiêu này phản ảnh lượng tăng (giảm) tương đối và thường được gọi là tốc độ tăng trưởng. Có ba loại tốc độ tăng (giảm) sau:

- *Tốc độ tăng (giảm) liên hoàn:*

$$a_i = \frac{y_i - y_{i-1}}{y_{i-1}} = t_i - 1 \qquad \text{Với i = 2, 3,…,n}$$

- *Tốc độ tăng (giảm) định gốc:*

$$b_i = \frac{y_i - y_1}{y_1} = T_i - 1 \qquad \text{Với i = 2, 3,…,n}$$

- *Tốc độ tăng (giảm) trung bình:* $\bar{a} = \bar{t} - 1$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Ví dụ, có dãy số thời gian về một doanh nghiệp:

Năm	09	10	11	12	13	14
Lợi nhuận (tỉ đ)	2	3	5	4	6	7

$$a_{10} = \frac{y_{10} - y_{09}}{y_{09}} = \frac{3-2}{2} = 0{,}5 = 50\%$$

$$a_{12} = \frac{y_{12} - y_{11}}{y_{11}} = \frac{4-5}{5} = -0{,}2 = -20\%$$

$$b_{11} = \frac{y_{11} - y_{09}}{y_{09}} = \frac{5-2}{2} = 1{,}5 = 150\%$$

$$b_{13} = \frac{y_{13} - y_{09}}{y_{09}} = \frac{6-2}{2} = 2 = 200\%$$

$$\bar{a} = \bar{t} - 1 = \sqrt[n-1]{T_n} - 1 = \sqrt[6-1]{\frac{y_{14}}{y_{09}}} - 1 = \sqrt[5]{\frac{7}{2}} - 1 = 0{,}28 = 28\%$$

Chú ý: *Nếu thứ tự thời gian của dữ liệu thời gian không được xét đến thì dữ liệu này cũng được mô tả bằng các số đo trung tâm và độ phân tán tương tự như dữ liệu chéo đã được trình bày ở chương 2.*

8.3. Biểu hiện xu hướng phát triển cơ bản của hiện tượng (làm nhẵn dãy số)

Để biểu hiện xu hướng phát triển chính của hiện tượng qua thời gian, người ta thường loại bỏ các biến động ngẫu nhiên, ngắn hạn ra khỏi dãy số. Phương pháp số trung bình trượt và phương pháp hồi quy thường được sử dụng.

8.3.1. Phương pháp số trung bình trượt

Phương pháp này thường được sử dụng để loại bỏ thành phần ngẫu nhiên hoặc cả thành phần ngẫu nhiên và thành phần thời vụ ra khỏi dãy số.

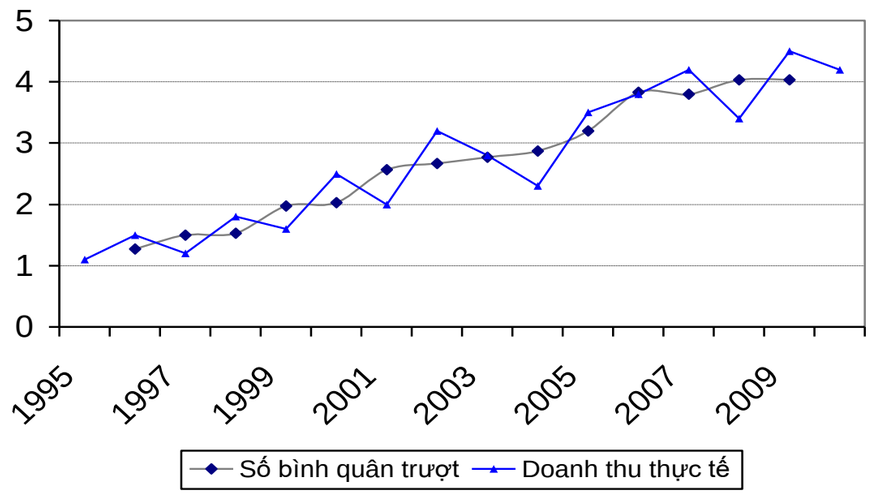
Mỗi số trung bình trượt được tính trung bình cộng từ nhóm k mức độ liên tiếp của dãy số. Số trung bình trượt thứ i được bắt đầu từ mức độ thứ i của dãy số. Số k được gọi là bước trượt hay khoảng trượt. *Người ta thường chọn bước trượt k với một số lưu ý sau:*

+ Nếu cần loại trừ thành phần ngẫu nhiên (trong dãy số không có thành phần thời vụ): Chọn k=3 với trường hợp dãy số biến động ngẫu nhiên không quá nhiều. Chọn k=5 hoặc lớn hơn với trường hợp dãy số có biến động ngẫu nhiên mạnh. Để tiện trình bày, nên chọn k là số lẽ.

+ Nếu cần loại trừ cả thành phần ngẫu nhiên và thành phần thời vụ (trong dãy số có thành phần thời vụ): Chọn k bằng số kỳ vụ trong năm.

Ví dụ, có dãy số thời gian về doanh thu của một doanh nghiệp qua 16 năm và số trung bình trượt với bước trượt k = 3 như sau:

Năm	Doanh thu (tỉ đồng)	Số trung bình trượt (k=3)
1995	1,1	-
1996	1,5	1,27
1997	1,2	1,50
1998	1,8	1,53
1999	1,6	1,97
2000	2,5	2,03
2001	2,0	2,57
2002	3,2	2,67
2003	2,8	2,77
2004	2,3	2,87
2005	3,5	3,20
2006	3,8	3,83
2007	4,2	3,80
2008	3,4	4,03
2009	4,5	4,03
2010	4,2	-



Đồ thị biểu diễn cho ta thấy rõ biến động thực tế và xu hướng biến động cơ bản của hiện tượng thông qua các số trung bình trượt.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Một cách tổng quát, giả sử có dãy số thời gian với các mức độ y_1, y_2, y_3, …, y_n. Số trung bình trượt tại thời gian t, được tính từ nhóm k = 2m+1 mức độ đối xứng với thời gian t là:

$$y_t^* = \frac{y_{t-m} + y_{t-m+1} + \ldots + y_t + \ldots + y_{t+m-1} + y_{t+m}}{2m+1}$$

Trong đó:

m = 1, 2, 3…

t = m+1, m+2, m+3,…, n-m

8.3.2. Phương pháp hồi quy xu thế

Phương pháp này thường được sử dụng để loại bỏ thành phần ngẫu nhiên và thành phần chu kỳ trong dãy số không có thành phần thời vụ để biểu hiện thành phần xu thế. Các dãy số thời gian đã được khử thành phần thời vụ hoặc các dãy số thời gian theo năm thích hợp với phương pháp này.

Nội dung của phương pháp là lựa chọn hàm số toán học xấp xỉ tốt nhất để biểu hiện xu thế phát triển của hiện tượng. Ở đây biến phụ thuộc là các mức độ của dãy số, biến độc lập là thứ tự thời gian của dãy số. Các dạng hàm có thể sử dụng là hàm tuyến tính, hàm mũ, hàm hypecbon, hàm paraboon. Tuy nhiên hàm tuyến tính là hàm thường được sử dụng nhất:

$$\hat{y}_t = b_o + b_1 t$$

Trong đó:

$\hat{y}_t$: Mức độ xu thế tại thời gian t

b_o: Hệ số chặn của hàm xu thế

b_1: Hệ số gốc của hàm xu thế

t: Thứ tự thời gian của dãy số

Các hệ số b_o và b_1 được xác định theo công thức (xem mục 7.1.3):

$$b_1 = \frac{\overline{y.t} - \bar{y}.\bar{t}}{\overline{t^2} - (\bar{t})^2} \quad \text{và} \quad b_o = \bar{y} - b_1.\bar{t}$$

Ví dụ, có dãy số thời gian về một doanh nghiệp:

Năm	13	14	15	16	17	18
Lợi nhuận (tỉ đ)	2	3	5	4	6	7

Dãy số này không có thành phần thời vụ. Có thể sử dụng phương pháp hồi qui để biểu hiện thành phần xu thế theo hàm tuyến tính như sau.

Năm	t	y_t	$y_t.t$	t^2
13	1	2	2	1
14	2	3	6	4
15	3	5	15	9
16	4	4	16	16
17	5	6	30	25
18	6	7	42	36
Trung bình	3,5	4,5	18,5	15,17

$$b_1 = \frac{\overline{y.t} - \bar{y}.\bar{t}}{\overline{t^2} - (\bar{t})^2} = \frac{18{,}5 - 4{,}5 \cdot 3{,}5}{15{,}17 - 3{,}5^2} = 0{,}94$$

$$b_o = \bar{y} - b_1.\bar{t} = 4{,}5 - 0{,}94 \cdot 3{,}5 = 1{,}2$$

Hàm xu thế lợi nhuận của doanh nghiệp: $\hat{y}_t = 1{,}2 + 0{,}94.t$

Sử dụng hàm xu thể ta có thể xác định lợi nhuận xu thế của doanh nghiệp từng năm sau khi loại bỏ thành phần ngẫu nhiên và chu kỳ (nếu có) như sau:

Năm	13	14	15	16	17	18
Lợi nhuận (tỉ đ)	2	3	5	4	6	7
Lợi nhuận xu thế (tỉ.đ)	2.14	3.09	4.03	4.97	5.91	6.86

8.4. Biểu hiện biến động thời vụ

Biến động thời vụ là những biến động có tính chu kỳ hàng năm. Nghĩa là hiện tượng có biểu hiện tăng lên hay giảm đi một cách rõ rệt vào những thời gian nhất định trong năm.

Để phân tích biểu hiện biến động thời vụ, người ta sử dụng phương pháp chỉ số thời vụ.

8.4.1. Chỉ số thời vụ giản đơn

Chỉ số thời vụ giản đơn thường được sử dụng cho các hiện tượng không có xu hướng phát triển rõ rệt.

Công thức tính chỉ số thời vụ giản đơn:

$$I_{S(i)} = \frac{\bar{y}_i}{\bar{y}} * 100 \quad (\%)$$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Trong đó:

$I_{S(i)}$: Chỉ số thời vụ kỳ vụ i

$\bar{y}_i$: Mức độ trung bình một kỳ vụ cùng tên qua các năm

$\bar{y}$: Mức độ trung bình một kỳ vụ nói chung qua các năm

Ở đây, *kỳ vụ là những khoảng thời gian bằng nhau trong năm mà hiện tượng có biểu hiện tăng giảm rõ rệt.* Người ta thường chọn kỳ vụ là tháng hay quí cho phù hợp với các nguồn số liệu trong thực tế.

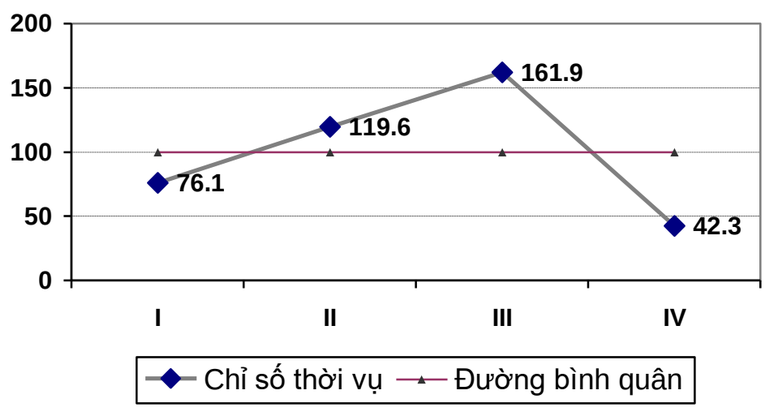
Để loại trừ biến động ngẫu nhiên, chỉ số thời vụ thường được tính từ dữ liệu khoảng ba năm trở lên.

Ví dụ, có chỉ số thời vụ về doanh số bán hàng của một doanh nghiệp như sau:

Quí	Doanh số bán (tỉ đồng)			Doanh số bán trung bình 1 quí	$I_{S(i)}$ (%)
	2016	2017	2018		
I	2,0	2,2	2,1	2,10	76,1
II	3,0	3,4	3,5	3,30	119,6
III	4,0	4,6	4,8	4,47	161,9
IV	1,0	1,2	1,3	1,17	42,3
Cộng	10,0	11,4	11,7	2,76	400,0

Chỉ số thời vụ giản đơn có nhược điểm là không loại trừ được thành phần biến động xu hướng ra khỏi biến động thời vụ. Do đó, ở đây biến động xu hướng, nếu có, cần được xem như là một thành phần đặc biệt của biến động thời vụ. Tuy nhiên, ưu điểm của nó là đơn giản nên dễ nhận thức. Nó phản ảnh mức độ dao động của hiện tượng quanh mức độ trung bình (một kỳ vụ) hàng năm.

Đồ thị biểu diễn chỉ số thời vụ giản đơn như sau:



8.4.2. Chỉ số thời vụ

Chỉ số thời vụ (tổng quát) loại trừ thành phần xu hướng ra khỏi biến động thời vụ. *Chỉ số này thường được sử dụng với các hiện tượng có xu hướng phát triển rõ rệt.*

Công thức tính chỉ số thời vụ (tổng quát):

$$I_{S(i)} = \frac{\sum_{j=1}^{n-1}\left(\frac{y_{ij}}{y_{ij}^{*}} * 100\right)}{n-1} \quad (\%)$$

Trong đó: y_{ij} : Mức độ thực tế kỳ vụ i năm j

y_{ij}^{*} : Mức độ xu hướng kỳ vụ i năm j

n : Số năm nghiên cứu

Mức độ xu hướng y_{ij}^{*} được xác định bằng hai lần trung bình trượt. Số trung bình trượt lần thứ nhất, với bước trượt k bằng số kỳ vụ trong năm. Số trung bình trượt lần thứ hai, với bước trượt k = 2 số trung bình trượt lần thứ nhất.

Chỉ số thời vụ (tổng quát) loại trừ biến động xu hướng nên chỉ phản ảnh riêng biến động thời vụ. Nó phản ánh mức độ dao động của hiện tượng quanh thành phần xu hướng. Tuy nhiên, nhược điểm của nó là phức tạp và trừu tượng hơn so với chỉ số thời vụ giản đơn.

Ví dụ, có doanh số bán hàng của một doanh nghiệp như sau:

Quí	Doanh số bán (tỉ đồng)			
	2015	2016	2017	2018
I	2,0	2,2	2,1	2,4
II	3,0	3,4	3,5	4,1
III	4,0	4,6	4,8	5,9
IV	1,0	1,2	1,3	1,5
Cộng	10,0	11,4	11,7	13,9

Việc tính chỉ số thời vụ được thực hiện như bảng sau.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Năm (j)	Quí (i)	Doanh số (y_{ij})	Trung bình trượt lần một	Trung bình trượt lần hai (y_{ij}^*)	$(y_{ij}/y_{ij}^*)*100$ (%)
2015	I	2,0	2,50	-	-
	II	3,0	2,55	-	-
	III	4,0	2,65	2,53	158,4
	IV	1,0	2,80	2,60	38,5
2016	I	2,2	2,85	2,73	80,7
	II	3,4	2,83	2,83	120,4
	III	4,6	2,85	2,84	162,1
	IV	1,2	2,90	2,84	42,3
2017	I	2,1	2,93	2,88	73,0
	II	3,5	3,00	2,91	120,2
	III	4,8	3,15	2,96	162,0
	IV	1,3	3,43	3,08	42,3
2018	I	2,4	3,48	3,29	73,0
	II	4,1		3,45	118,8
	III	5,9		-	-
	IV	1,5		-	-

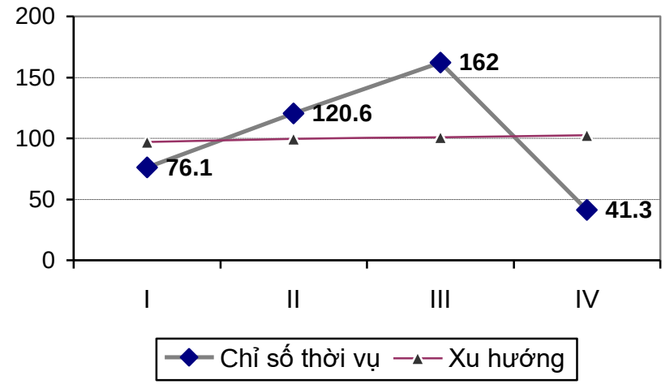
Từ cột cuối ta lập bảng sau:

Quý (i)	$(y_{ij}/y_{ij}^*)*100$ (%)				$I_{S(i)}$ (%)	$I_{S(i)(đc)}$ (%)
	2015	2016	2017	2018		
I	-	80,7	73,0	73,0	75,6	76,1
II	-	120,4	120,2	118,8	119,8	120,6
III	158,4	162,1	162,0	-	160,9	162,0
IV	38,5	42,3	42,3	-	41,0	41,3
Cộng	-	-	-	-	397,2	400,0

Các chỉ số thời vụ thường được điều chỉnh sao cho tổng của chúng bằng 100% nhân với số kỳ vụ trong năm. Do đó, $I_{S(i)(đc)} = I_{S(i)}$* H với H là hệ số điều chỉnh:

$$H = 400/397,2 = 1,007$$

Chỉ số thời vụ được trình bày trên đồ thị như sau:



Chú ý: * Chỉ số thời vụ trên cũng có thể được tính theo công thức:

$$I_{S(i)} = Median(\frac{y_{ij}}{y_{ij}^*}*100) \quad \%$$

8.5. Phân tích các thành phần của dãy số thời gian

Các mức độ của một dãy số thời gian có thể được phân tích dựa trên mô hình nhân từ các thành phần hợp thành: xu thế, chu kỳ, thời vụ, và ngẫu nhiên.

$$y_t = y_{(T.C.S.I)} = T.C.S.I$$

Trong đó:

T: Thành phần xu thế
C: Thành phần chu kỳ
S: Thành phần thời vụ
I: Thành phần ngẫu nhiên

- Trích xuất thành phần thời vụ (S): Thành phần thời vụ được trích xuất trước tiên thông qua chỉ số thời vụ (I_S) xác định như trong mục 8.4.2.

- Trích xuất thành phần xu thế (T): Thành phần xu thế được trích xuất qua hai bước.

+ Bước 1: Trích xuất dãy số không có thành phần thời vụ $y_{(T.C.I)}$ theo công thức: $y_{(T.C.I)} = y_t / I_S$. Trong đó I_S là chỉ số thời vụ.

+ Bước 2: Sử dụng phương pháp hồi qui như đã trình bày ở mục 8.3.2. để trích xuất thành phần xu thế $y_{(T)}$ từ dãy số đã loại thành thần thời vụ $y_{(T.C.I)}$.

- Trích xuất thành phần chu kỳ (C): Thành phần chu kỳ được trích xuất qua hai bước.

+ Bước 1: Trích xuất dãy số gồm hai thành phần chu kỳ và ngẫu nhiên $y_{(C.I)}$ theo công thức: $y_{(T.C.I)} = y_{(T.C.I)} / y_{(T)}$.

+ Bước 2: Thành phần chu kỳ thể hiện qua chỉ số chu kỳ $I_{(C)}$ được trích xuất bằng số trung bình trượt của dãy số $y_{(C.I)}$ với bước trượt 3 hoặc 5 mức độ.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

- Trích xuất thành phần ngẫu nhiên (I): Thành phần ngẫu nhiên được trích xuất qua chỉ số ngẫu nhiên $I_{(I)}$ theo công thức: $I_{(I)} = y_{(C.I)} / I_{(C)}$.

Ví dụ: Việc trích xuất các thành phần của dãy số thời gian trong ví dụ ở mục 8.4.2 được thể hiện qua bảng sau.

Năm (j)	Quí (i)	t	**Doanh số y_t**	**Thời vụ I_S (%)**	$y_{(T.C.I)}$	**Xu thế $y_{(T)}$**	$y_{(C.I)}$	**Chu kỳ $I_{(C)}$ (%)**	**Ngẫu nhiên $I_{(I)}$ (%)**
1	2	3	**4**	**5**	6=4/5	**7**	8= 6/7	**9**	**10**=8/9
2015	I	1	2,0	76,1	2,63	2,40	1,10	-	-
	II	2	3,0	120,6	2,49	2,47	1,01	103	98,3
	III	3	4,0	162,0	2,47	2,54	0,97	97	100,3
	IV	4	1,0	41,3	2,42	2,61	0,93	99	93,5
2016	I	5	2,2	76,1	2,89	2,69	1,07	101	106,6
	II	6	3,4	120,6	2,82	2,76	1,02	103	98,9
	III	7	4,6	162,0	2,84	2,83	1,00	101	99,4
	IV	8	1,2	41,3	2,91	2,90	1,00	98	102,6
2017	I	9	2,1	76,1	2,76	2,98	0,93	96	96,5
	II	10	3,5	120,6	2,90	3,05	0,95	94	101,0
	III	11	4,8	162,0	2,96	3,12	0,95	96	98,7
	IV	12	1,3	41,3	3,15	3,20	0,98	97	102,0
2018	I	13	2,4	76,1	3,15	3,27	0,96	99	97,5
	II	14	4,1	120,6	3,40	3,34	1,02	102	100,2
	III	15	5,9	162,0	3,64	3,41	1,07	104	102,5
	IV	16	1,5	41,3	3,63	3,49	1,04	-	-

Mô hình nhân: (4) = (7) * (9) * (5) * (10)

Trong đó:

- Cột (4): Dãy số thực tế
- Cột (5): Thành phần thời vụ I_S, được lấy từ chỉ số thời vụ tính như ở mục 8.4.2.
- Cột (6): Dãy số không có thành phần thời vụ $y_{(T.C.I)}$ được xác định bằng cách lấy cột (4) chia cho cột (5).
- Cột (7): Thành phần xu thế $y_{(T)}$ được xác định thông qua hàm xu thế $y_{(T)} = \hat{y}_t$. Hàm xu thế $\hat{y}_t = 2{,}32 + 0{,}07.t$ được xác định bằng phương pháp hồi qui từ cột (5) và cột (3) theo các công thức như ở mục 8.3.2 (có thể thực hiện trên EXCEL hay SPSS).
- Cột (8): Dãy số gồm hai thành phần chu kỳ và ngẫu nhiên $y_{(C.I)}$ được xác định bằng cách lấy cột (6) chia cho cột (7).
- Cột (9): Thành phần chu kỳ $I_{(C)}$ được xác định bằng trung bình trượt với bước trượt 3 mức độ từ cột (8).
- Cột (10): Thành phần ngẫu nhiên $I_{(I)}$ được xác định bằng cách lấy cột (8) chia cho cột (9).

8.6. Một số phương pháp dự đoán dựa trên dãy số thời gian

Các dự đoán thống kê nói chung dựa trên một giả định căn bản, đó là sự biến động trong tương lai của hiện tượng nói chung sẽ giống với sự biến động của hiện tượng trong quá khứ. Nói cách khác, các yếu tố tác động đến hiện tượng trong quá khứ được xem là sẽ vẫn tiếp tục tác động gần tương tự như vậy đến hiện tượng trong tương lai.

Giả định nói trên là khó có được trong thực tế đầy biến động của môi trường kinh tế-xã hội. Tuy nhiên, các con số dự đoán như vậy cũng cung cấp được những thông tin trợ giúp rất hữu ích cho các nhà quản lý.

8.6.1. Phương pháp dự đoán dựa trên lượng tăng (giảm) tuyệt đối trung bình

Công thức dự đoán: $\hat{y}_{n+L} = y_n + \bar{\partial} * L$

Trong đó: $\hat{y}_{n+L}$: Trị số dự đoán tại thời gian n+L

y_n : Trị số tại thời gian cuối của dãy số

$\bar{\partial}$: Lượng tăng (giảm) tuyệt đối trung bình

L : Tầm xa dự đoán

Ví dụ, có dãy số thời gian về doanh số của một doanh nghiệp:

Năm	2014	2015	2016	2017	2018
Doanh số (tỉ đồng)	1,20	1,65	1,97	2,32	2,82

$$\bar{\partial} = \frac{\Delta_n}{n-1} = \frac{2{,}82 - 1{,}20}{5-1} = 0{,}405$$

Doanh số dự đoán cho năm 2019 (L=1):

$$\hat{y}_{2019} = 2{,}82 + 0{,}405 * 1 = 3{,}225 \text{ (tỉ đồng)}$$

Phương pháp này được dùng khi lượng tăng (giảm) tuyệt đối liên hoàn xấp xỉ nhau qua các thời gian. Phương pháp đơn giản, dễ hiểu. Nhược điểm của nó là chịu ảnh hưởng mạnh của biến động ngẫu nhiên trong mức độ cuối cùng của dãy số.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

8.6.2. Phương pháp dự đoán dựa trên tốc độ phát triển trung bình

Công thức dự đoán: $\hat{y}_{n+L} = y_n * \bar{t}^L$

Trong đó: $\hat{y}_{n+L}$: Trị số dự đoán tại thời gian n+L

y_n : Trị số tại thời gian cuối của dãy số

$\bar{t}$: Tốc độ phát triển trung bình

L : Tầm xa dự đoán

Ví dụ, có dãy số thời gian về doanh số của một doanh nghiệp:

Năm	2014	2015	2016	2017	2018
Doanh số (tỉ đồng)	1,20	1,65	2,17	2,82	3,42

$$\bar{t} = \sqrt[n-1]{T_n} = \sqrt[5-1]{\frac{3,42}{1,20}} = 1,299$$

Doanh số dự đoán cho năm 2020 (L=2):

$$\hat{y}_{2020} = 3,42 * 1,299^2 = 5,77 \text{ (tỉ đồng)}$$

Phương pháp này thường được dùng khi tốc độ phát triển liên hoàn hàng năm xấp xỉ nhau. Phương pháp đơn giản, dễ hiểu. Nhược điểm của nó là chịu ảnh hưởng mạnh của biến động ngẫu nhiên trong mức độ cuối cùng của dãy số.

8.6.3. Phương pháp san bằng mũ đơn giản (simple exponential smoothing)

Công thức dự đoán: $\hat{y}_{n+L} = \bar{y}_n$

Trong đó: $\hat{y}_{n+L}$: Trị số dự đoán tại thời gian n+L

$\bar{y}_n$: Trị số san bằng tại thời gian cuối của dãy số

L : Tầm xa dự đoán

Trị số san bằng tại thời gian cuối của dãy số được san bằng cập nhật từ đầu dãy số theo công thức:

$$\bar{y}_t = \alpha y_t + (1-\alpha)\bar{y}_{t-1}$$

Trong đó: α : Trọng số (hằng số san bằng) của y_t

1- α : Trọng số (hằng số san bằng) của $\bar{y}_{t-1}$

y_t : Trị số thực tế tại thời gian t

Hằng số α lấy giá trị trong khoảng (0,1) được chọn dựa vào một trong các căn cứ sau:

+ Dựa vào kinh nghiệm và sự phán đoán của người dự đoán

+ Nếu đồ thị của hiện tượng có nhiều biến động ngẫu nhiên nên chọn giá trị lớn và ngược lại

+ Thử các giá trị 0,2 ; 0,4 ; 0,6 ; 0,8 để chọn giá trị phản ảnh tốt nhất xu hướng biến động của hiện tượng

+ Chọn giá trị α sao cho tổng bình phương sai số dự đoán sau là nhỏ nhất:

$$SS = \sum_{t=2}^{n}(y_t - \hat{y}_t)^2$$

Trị số san bằng đầu tiên thường được chọn là: $\bar{y}_1 = y_1$

Ví dụ, có dãy số thời gian về doanh số của một doanh nghiệp:

Năm	2013	2014	2015	2016	2017	2018
Doanh số (tỉ đồng)	0,96	1,20	1,65	1,97	2,32	2,28

Chọn α = 0,7 , $\bar{y}_{13} = y_{13} = 0,96$

Dựa vào công thức san bằng: $\bar{y}_t = \alpha y_t + (1-\alpha)\bar{y}_{t-1}$

$\bar{y}_{14} = \alpha y_{14} + (1-\alpha)\bar{y}_{13}$

= 0,7*1,20 + (1-0,7)*0,96 = 1,13

$\bar{y}_{15} = \alpha y_{15} + (1-\alpha)\bar{y}_{14}$

= 0,7*1,65 + (1-0,7)*1,13 = 1,49

Tương tự cho các năm 16, 17 và 18, Ta có bảng sau:

Năm	y_t	$\bar{y}_t$	$\hat{y}_t = \bar{y}_{t-1}$	$(y_t - \hat{y}_t)^2$
2013	0,96	0,96	-	-
2014	1,20	1,13	0,96	0,06
2015	1,65	1,49	1,13	0,27
2016	1,97	1,83	1,49	0,23
2017	2,32	2,17	1.83	0,24
2018	2,28	2,25	2,17	0,01
Tổng (sai số dự đoán)				0,81

Doanh số dự đoán cho năm 2019 (L=1):

$$\hat{y}_{2019} = \bar{y}_{2018} = 2,25 \text{ (tỉ đồng)}$$

Doanh số dự đoán cho năm 2020 (L=2):

$$\hat{y}_{2020} = \bar{y}_{2018} = 2,25 \text{ (tỉ đồng)}$$

Sai số dự đoán: $SS = \sum_{t=2}^{n}(y_t - \hat{y}_t)^2 = 0,81$

Phương pháp này thường được sử dụng với các hiện tượng có nhiều biến động ngẫu nhiên nhưng không có xu hướng phát triển rõ rệt.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

8.6.4. Phương pháp Holt-Winters với hiện tượng không có tính thời vụ

Công thức dự đoán: $\hat{y}_{n+L} = \bar{y}_n + \bar{\partial}_n \cdot L$

Trong đó:

$\hat{y}_{n+L}$: Trị số dự đoán tại thời gian n+L

$\bar{y}_n$: Trị số san bằng tại thời gian cuối của dãy số

$\bar{\partial}_n$: Lượng tăng (giảm) liên hoàn san bằng tại thời gian cuối

L : Tầm xa dự đoán

Trị số san bằng và lượng tăng (giảm) liên hoàn san bằng tại thời gian cuối của dãy số được san bằng cập nhật từ đầu dãy số theo các công thức:

$$\bar{y}_t = \alpha y_t + (1-\alpha)(\bar{y}_{t-1} + \bar{\partial}_{t-1})$$

$$\bar{\partial}_t = \beta(\bar{y}_t - \bar{y}_{t-1}) + (1-\beta)\bar{\partial}_{t-1}$$

Trong đó: α : Trọng số (hằng số san bằng) của y_t

$1-\alpha$: Trọng số (hằng số san bằng) của $(\bar{y}_{t-1} + \bar{\partial}_{t-1})$

β : Trọng số (hằng số san bằng) của $(\bar{y}_t - \bar{y}_{t-1})$

$1-\beta$: Trọng số (hằng số san bằng) của $\bar{\partial}_{t-1}$

y_t : Trị số thực tế tại thời gian t

Các trị số san bằng đầu tiên thường được chọn là: $\bar{y}_2 = y_2$, $\bar{\partial}_2 = y_2 - y_1$.

Các hằng số α , β lấy giá trị trong khoảng (0,1) được chọn dựa vào một trong các căn cứ sau:

+ Dựa vào kinh nghiệm và sự phán đoán của người dự đoán;

+ Nếu hiện tượng có nhiều biến động ngẫu nhiên nên chọn giá trị lớn và ngược lại;

+ Thử các giá trị từ 0,1 đến 0,4 để chọn giá trị phản ảnh tốt nhất xu hướng biến động của hiện tượng;

+ Chọn giá trị α và β sao cho tổng bình phương các sai số dự đoán sau là nhỏ nhất:

$$SS = \sum_{t=2}^{n} (y_t - \hat{y}_t)^2$$

Ví dụ, có dãy số thời gian về doanh số của một doanh nghiệp:

Năm	2013	2014	2015	2016	2017	2018
Doanh số (tỉ đồng)	0,96	1,20	1,65	1,97	2,32	2,28

Nếu chọn : $\alpha = 0,7$, $\beta = 0,6$

$\bar{y}_{14} = y_{14} = 1,20$, $\bar{\partial}_{14} = y_{14} - y_{13} = 1,20 - 0,96 = 0,24$

Dựa vào công thức san bằng: $\bar{y}_t = \alpha y_t + (1-\alpha)(\bar{y}_{t-1} + \bar{\partial}_{t-1})$

$$\bar{\partial}_t = \beta(\bar{y}_t - \bar{y}_{t-1}) + (1-\beta)\bar{\partial}_{t-1}$$

Ta được:

$$\bar{y}_{15} = \alpha y_{15} + (1-\alpha)(\bar{y}_{14} + \bar{\partial}_{14})$$

$= 0,7*1,65 + (1 - 0,7)(1,20+0,24) = 1,59$

$$\bar{\partial}_{15} = \beta(\bar{y}_{15} - \bar{y}_{14}) + (1-\beta)\bar{\partial}_{14}$$

$= 0,6*(1,59-1,20) + (1 - 0,6)0,24 = 0,33$

$$\bar{y}_{16} = \alpha y_{16} + (1-\alpha)(\bar{y}_{15} + \bar{\partial}_{15})$$

$= 0,7*1,97 + (1 - 0,7)(1,59+0,33) = 1,95$

$$\bar{\partial}_{16} = \beta(\bar{y}_{16} - \bar{y}_{15}) + (1-\beta)\bar{\partial}_{15}$$

$= 0,6* (1,95-1,59) + (1 - 0,6)0,33 = 0,35$

Tương tự cho các năm 17 và 18, ta có bảng sau:

Năm	y_t	$\bar{y}_t$	$\bar{\partial}_t$	$\hat{y}_t = \bar{y}_{t-1} + \bar{\partial}_{t-1}$	$(y_t - \hat{y}_t)^2$
2013	0,96	-	-	-	-
2014	1,20	1,20	0,24	-	-
2015	1,65	1,59	0,33	1,44	0,0410
2016	1,97	1,95	0,35	1,92	0,0025
2017	2,32	2,32	0,36	2,30	0,0004
2018	2,28	2,40	0,19	2,68	0,1600
Tổng	-	-	-	-	0,2070

Doanh số dự đoán cho năm 2019 (L=1):

$\hat{y}_{2019} = \bar{y}_{2018} + \bar{\partial}_{2018} * 1 = 2,40 + 0,19 * 1 = 2,59$ (tỉ đồng)

Sai số dự đoán: $SS = \sum_{t=2}^{n} (y_t - \hat{y}_t)^2 = 0,207$

Phương pháp này thường dùng với hiện tượng có xu hướng tăng giảm rõ rệt nhưng không có tính thời vụ.

8.6.5. Phương pháp ngoại suy hàm xu thế

Công thức dự đoán: $\hat{y}_{n+L} = f(n+L)$

Trong đó: $\hat{y}_{n+L}$: Trị số dự đoán tại thời gian n+L

$\hat{y}_t = f(t)$ *: Hàm hồi qui xu thế được xác định dựa vào dãy số thời gian.*

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Ví dụ, có dãy số thời gian về một doanh nghiệp:

Năm	13	14	15	16	17	18
Doanh số (tỉ đ)	2	3	5	4	6	7

Hàm hồi qui xu thế từ dãy số: $\hat{y}_t = 1{,}2 + 0{,}94.t$ (xem mục 8.3.2)

Doanh số dự đoán năm 2019 (t=7): $\hat{y}_{2019} = 1{,}2 + 0{,}94 * 7 = 7{,}78$ (tỉ.đ)

8.6.6. Phương pháp dự đoán dựa trên mô hình nhân cho các hiện tượng có tính thời vụ

Công thức dự đoán tổng quát: $\hat{y}_{n+L} = \hat{T} \cdot \hat{C} \cdot \hat{S}$

Trong đó: $\hat{y}_{n+L}$: Trị số dự đoán tại thời gian n+L

$\hat{T}$: Trị số dự đoán của thành phần xu thế

$\hat{C}$: Trị số dự đoán của thành phần chu kỳ

$\hat{S}$: Trị số dự đoán của thành phần thời vụ

Việc dự đoán có thể tiến hành dựa trên bảng phân tích các thành phần của dãy số thời gian như trong mục 8.6.

Năm (j)	Quí (i)	t	**Doanh số y_t**	**Thời vụ I_S (%)**	$y_{(T.C.I)}$	**Xu thế $y_{(T)}$**	$y_{(C.I)}$	**Chu kỳ $I_{(C)}$ (%)**
1	2	3	**4**	**5**	6=4/5	**7**	8= 6/7	**9**
2015	I	1	2,0	76,1	2,63	2,40	1,10	-
	II	2	3,0	120,6	2,49	2,47	1,01	103
	III	3	4,0	162,0	2,47	2,54	0,97	97
	IV	4	1,0	41,3	2,42	2,61	0,93	99
2016	I	5	2,2	76,1	2,89	2,69	1,07	101
	II	6	3,4	120,6	2,82	2,76	1,02	103
	III	7	4,6	162,0	2,84	2,83	1,00	101
	IV	8	1,2	41,3	2,91	2,90	1,00	98
2017	I	9	2,1	76,1	2,76	2,98	0,93	96
	II	10	3,5	120,6	2,90	3,05	0,95	94
	III	11	4,8	162,0	2,96	3,12	0,95	96
	IV	12	1,3	41,3	3,15	3,20	0,98	97
2018	I	13	2,4	76,1	3,15	3,27	0,96	99
	II	14	4,1	120,6	3,40	3,34	1,02	102
	III	15	5,9	162,0	3,64	3,41	1,07	104
	IV	16	1,5	41,3	3,63	3,49	1,04	-

Việc dự đoán theo công thức trên đòi hỏi dự đoán từng thành phần trong công thức rồi nhân chúng lại với nhau để được kết quả dự đoán cuối cùng.

Ví dụ với bảng trên, dự đoán doanh số quí II năm 2019 hay t =18:

- Thành phần xu thế: Dự đoán theo hàm xu thế (xem mục 8.6.5) được xác định dựa vào cột (6) và cột (3) $\hat{y}_t = 2{,}32 + 0{,}07.t$.

$$\hat{T} = \hat{y}_{18} = 2{,}32 + 0{,}07.18 = 3{,}58$$

- Thành phần chu kỳ: Dựa vào chỉ số chu kỳ quí II. Thành phần này thường rất khó dự đoán. Tuy nhiên nhìn trên bảng, ta thấy chỉ số chu kỳ biến động không nhiều nên có thể lấy chỉ số dự đoán bằng 1.

$$\hat{C} = I_{C(II)} = 1$$

- Thành phần thời vụ: Dựa vào chỉ số thời vụ quí II.

$$\hat{S} = I_{S(II)} = 120{,}6\%$$

Doanh số dự đoán của quí II năm 2019:

$$\hat{y}_{II/2019} = \hat{T} * \hat{C} * \hat{S} = 3{,}58 * 1 * 1{,}206 = 4{,}32 \quad \textit{(tỉ.đ)}$$

Chú ý: Để khắc phục hạn chế của việc dự đoán riêng thành phần chu kỳ như nói trên, có thể dự đoán chung cho cả thành phần xu thế và thành phần chu kỳ bằng phương pháp Holt-Winters dựa trên dãy số $y_{(T.C.I)}$ trong cột (6) của bảng trên. Sau đó chỉ cần nhân kết quả dự đoán này với chỉ số thời vụ tương ứng.

8.6.7. Các tiêu chuẩn lựa chọn phương pháp dự đoán

Việc lựa chọn phương pháp dự đoán tốt nhất nên dựa vào kinh nghiệm và bản chất của hiện tượng. Tuy nhiên trong trường hợp chưa có kinh nghiệm dự đoán với một hiện tượng nào đó, có thể dựa vào một trong các tiêu chuẩn sau để lựa chọn.

1) Số trung bình các sai số dự đoán tuyệt đối:

$$MAE = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{n}$$

n là số các cặp y_t và $\hat{y}_t$ trên dãy số thời gian.

2) Số trung bình bình phương các sai số dự đoán:

$$MSE = \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}$$

n là số các cặp y_t và $\hat{y}_t$ trên dãy số thời gian.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

8.7. Chỉ số

8.7.1. Chỉ số giá cá thể

Chỉ số giá cá thể phản ảnh biến động giá của từng mặt hàng qua thời gian. Chỉ số giá cá thể còn được gọi là giá tương đối.

Công thức tính: $i_p = \frac{p_1}{p_0} \cdot 100$

Trong đó:

P_1 : Đơn giá của mặt hàng ở kỳ nghiên cứu

P_0 : Đơn giá của mặt hàng ở kỳ gốc

Kỳ gốc làm căn cứ so sánh có thể được chọn liên hoàn hay cố định tuỳ theo yêu cầu thông tin và nguồn dữ liệu.

Ví dụ, có đơn giá một mặt hàng như sau:

Năm	2005	2006	2007	2008	2009	2010
Đơn giá (nghìn đồng/kg)	25	27	28	32	33	35

Ta có bảng các chỉ số giá cá thể sau:

Năm	2005	2006	2007	2008	2009	2010
Chỉ số giá định gốc (%)	100	108	112	128	132	140
Chỉ số giá liên hoàn (%)	-	108,0	103,7	114,3	103,1	106,1

8.7.2. Chỉ số giá tổng hợp

Chỉ số giá tổng hợp phản ảnh biến động giá của từng nhóm hàng hóa hay toàn bộ hàng hóa trên một thị trường qua thời gian.

a. Chỉ số giá tổng hợp không có trọng số:

$$I_p = \frac{\sum p_{i1}}{\sum p_{i0}} \cdot 100$$

Trong đó:

P_{i1} : Đơn giá của mặt hàng i ở kỳ nghiên cứu

P_{i0} : Đơn giá của mặt hàng i ở kỳ gốc

Chỉ số giá tổng hợp không có trọng số có ưu điểm là đơn giản. Tuy nhiên, chỉ số này nhạy cảm với đơn vị tính của đơn giá nên ít được sử dụng trong thực tế.

b. Chỉ số giá tổng hợp có trọng số:

Triết lý cơ sở của chỉ số giá tổng hợp có trọng số là mỗi mặt hàng trong nhóm hàng cần có trọng số tương ứng với tầm quan trọng của nó. Trong hầu hết các trường hợp, lượng tiêu thụ của từng mặt hàng là thước đo tầm quan trọng tốt nhất vì vậy nó thường được dùng làm trọng số của chỉ số giá tổng hợp có trọng số.

- Công thức tổng quát của chỉ số giá tổng hợp có trọng số:

$$I_p = \frac{\sum p_{i1} q_i}{\sum p_{i0} q_i} \cdot 100$$

Trong đó:

p_{i1} : Đơn giá của mặt hàng i ở kỳ nghiên cứu

P_{i0} : Đơn giá của mặt hàng i ở kỳ gốc

q_i : Trọng số (lượng tiêu thụ của mặt hàng i từ một nguồn nào đó)

- Chỉ số giá tổng hợp Laspeyres: *Sử dụng trọng số là lượng tiêu thụ ở kỳ gốc.*

$$I_p = \frac{\sum p_{i1} q_{i0}}{\sum p_{i0} q_{i0}} \cdot 100$$

Trong đó:

p_{i1} : Đơn giá của mặt hàng i ở kỳ nghiên cứu

P_{i0} : Đơn giá của mặt hàng i ở kỳ gốc

q_{i0} : Trọng số (lượng tiêu thụ mặt hàng i kỳ gốc)

- Chỉ số giá tổng hợp Paasche: *Sử dụng trọng số là lượng tiêu thụ ở kỳ nghiên cứu.*

$$I_p = \frac{\sum p_{i1} q_{i1}}{\sum p_{i0} q_{i1}} \cdot 100$$

Trong đó:

p_{i1} : Đơn giá của mặt hàng i ở kỳ nghiên cứu

P_{i0} : Đơn giá của mặt hàng i ở kỳ gốc

q_{i1} : Trọng số (lượng tiêu thụ mặt hàng i kỳ gốc)

Ví dụ, có dữ liệu về tình hình tiêu thụ 4 mặt hàng như sau:

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Mặt hàng	Đơn vị tính	Đơn giá (1000đ)		Lượng tiêu thụ	
		2015	2016	2015	2016
Cá thu	Kg	70	75	4000	6000
Thịt bò	Kg	75	80	10000	12000
Dầu ăn	Lít	10	12	2000	2200
Mở heo	Kg	8	9	800	600

Chỉ số giá tổng hợp Laspeyres:

$$I_p = \frac{\sum p_{i1}q_{i0}}{\sum p_{i0}q_{i0}} = \frac{75\cdot 4000 + 80\cdot 10000 + 12\cdot 2000 + 9\cdot 800}{70\cdot 4000 + 75\cdot 10000 + 10\cdot 2000 + 8\cdot 800}100 = 107{,}1\%$$

Chỉ số giá tổng hợp Paasche:

$$I_p = \frac{\sum p_{i1}q_{i1}}{\sum p_{i0}q_{i1}} = \frac{75\cdot 6000 + 80\cdot 12000 + 12\cdot 2200 + 9\cdot 600}{70\cdot 6000 + 75\cdot 12000 + 10\cdot 2200 + 8\cdot 600}100 = 107{,}6\%$$

- Tính chỉ số giá tổng hợp từ các chỉ số giá cá thể:

+ Công thức tổng quát: $I_p = \frac{\sum i_p(p_{i0}q_i)}{\sum p_{i0}q_i}\cdot 100$

Trong đó:

i_p : Chỉ số giá cá thể mặt hàng i

$p_{i0}q_i$: Trọng số

Khi sử dụng q_i là lượng tiêu thụ ở kỳ gốc, *trọng số sẽ là doanh số tiêu thụ ở kỳ gốc.* Công thức trên trở thành:

+ Công thức dạng Laspeyres: $I_p = \frac{\sum i_p(p_{i0}q_{i0})}{\sum p_{i0}q_{i0}}\cdot 100$

Khi sử dụng q_i trong công thức là lượng tiêu thụ của mặt hàng i ở kỳ nghiên cứu, *trọng số sẽ là doanh số tiêu thụ ở nghiên cứu theo giá kỳ gốc.* Công thức trên trở thành:

+ Công thức xấp xỉ dạng Paasche: $I_p = \frac{\sum i_p(p_{i0}q_{i1})}{\sum p_{i0}q_{i1}}\cdot 100$

Ví dụ: Có dữ liệu về tình hình tiêu thụ 4 mặt hàng như sau:

Mặt hàng	Đơn vị tính	Đơn giá 2010 (1000đ)	Lượng tiêu thụ 2016	Chỉ số giá cá thể 2016/2010 i_p (%)
Cá thu	Kg	70	6000	107,1
Thịt bò	Kg	75	12000	106,7
Dầu ăn	Lít	10	2200	120,0
Mở heo	Kg	8	600	112,5

Chỉ số giá tổng hợp:

$$I_p = \frac{\sum i_p(p_{i0}q_{i1})}{\sum p_{i0}q_{i1}}\cdot 100 = \frac{107{,}1(70\cdot 6000) + 106{,}7(75\cdot 12000) + \ldots + 112{,}5(8\cdot 600)}{70\cdot 6000 + 75\cdot 12000 + \ldots + 8\cdot 600} = 107{,}3\%$$

- Một số vấn đề khi tính chỉ số giá tổng hợp:

+ Vấn đề chọn các mặt hàng: Khi tính chỉ số giá tổng hợp cho một thị trường nào đó, thông thường số mặt hàng thường rất lớn. Không thể tính chỉ số giá cho tất cả các mặt hàng này. Thay vào đó, một mẫu các mặt hàng đại diện được sử dụng. Chẳng hạn, khi tính chỉ số giá tiêu dùng (CPI), số mặt hàng có thể được xem là có liên quan đến người tiêu dùng có thể lên đến 2000 hay hơn. Tuy nhiên, chỉ khoảng 400 mặt hàng được chọn để tính CPI. Một cuộc khảo sát các kiểu mua hàng của người tiêu dùng và một quá trình đánh giá kỹ lưỡng được tiến hành để chọn ra các mặt hàng đại diện.

+ Vấn đề chọn kỳ gốc: Thông thường chỉ số giá được tính cho một dãy số thời gian để phục vụ việc nghiên cứu các đặc điểm biến động của nó. Việc chọn thời kỳ nào làm kỳ gốc phải dựa trên sự phán xét của người xây dựng chỉ số. Một qui tắc hướng dẫn chung là kỳ gốc không nên quá xa so với thời kỳ nghiên cứu. Vì vậy, kỳ gốc phải được điều chỉnh định kỳ để gần hơn với kỳ nghiên cứu. Chẳng hạn CPI của Mỹ chọn kỳ gốc năm 1967 sau đó được điều chỉnh sang năm 1982.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

+ Vấn đề thay đổi chất lượng hàng hóa: Một vấn đề nảy sinh khi tính chỉ số giá là sự thay đổi chất lượng sản phẩm từ thời kỳ này sang thời kỳ khác. Điều này dẫn đến ý nghĩa của chỉ số giá không còn chính xác. Tuy nhiên việc khử ảnh hưởng thay đổi chất lượng ra khỏi chỉ số giá là rất khó khăn. Vì vậy chỉ những thay đổi chất lượng rõ rệt về chất lượng sản phẩm mới được xem xét. Khi đó những sản phẩm này phải được xem xét loại ra khỏi mẫu đại diện để tính chỉ số.

8.7.3. Chỉ số khối lượng cá thể: phản ảnh biến động khối lượng sản xuất hay tiêu thụ của từng mặt hàng qua thời gian.

Công thức tính: $i_q = \frac{q_1}{q_0} \cdot 100$

Trong đó:

q_1 : lượng sản xuất hay tiêu thụ của mặt hàng ở kỳ nghiên cứu

q_0 : lượng sản xuất hay tiêu thụ của mặt hàng ở kỳ gốc

8.7.4. Chỉ số khối lượng tổng hợp: *phản ảnh biến động khối lượng sản xuất hay tiêu thụ của nhóm nhiều sản phẩm qua thời gian.*

+ Công thức tổng quát: $I_q = \frac{\sum q_{i1} w_i}{\sum q_{i0} w_i} \cdot 100$

Trong đó:

q_{i1} : lượng sản xuất hay tiêu thụ sản phẩm i ở kỳ nghiên cứu

q_{i0} : lượng sản xuất hay tiêu thụ sản phẩm i ở kỳ gốc

w_i : Trọng số

Trọng số w_i khi tính chỉ số khối lượng sản xuất thường được chọn là giá cố định hay giá trị tăng thêm của sản phẩm. Giá trị tăng thêm của một loại sản phẩm bằng giá trị bán ra trừ đi chi phí đầu vào.

+ Chỉ số khối lượng tổng hợp với trọng số là giá kỳ gốc:

$$I_q = \frac{\sum q_{i1} p_{i0}}{\sum q_{i0} p_{i0}} \cdot 100$$

Trong đó: q_{i1} : lượng sản xuất hay tiêu thụ mặt hàng i ở kỳ nghiên cứu

q_{i0} : lượng sản xuất hay tiêu thụ mặt hàng i ở kỳ gốc

P_{i0} : Trọng số (Đơn giá của mặt hàng i ở kỳ gốc)

Ví dụ, có dữ liệu về tình hình tiêu thụ 4 mặt hàng như sau:

Mặt hàng	Đơn vị tính	Đơn giá (1000đ)		Lượng tiêu thụ	
		2014	2016	2014	2016
Cá thu	Kg	70	75	4000	6000
Thịt bò	Kg	75	80	10000	12000
Dầu ăn	Lít	10	12	2000	2200
Mở heo	Kg	8	9	800	600

Chỉ số khối lượng tổng hợp:

$$I_q = \frac{\sum q_{i1} p_{i0}}{\sum q_{i0} p_{i0}} \cdot 100$$

$$= \frac{6000 \cdot 70 + 12000 \cdot 75 + 2200 \cdot 10 + 600 \cdot 8}{4000 \cdot 70 + 10000 \cdot 75 + 2000 \cdot 10 + 800 \cdot 8} 100 = 127{,}5\%$$

+ Chỉ số khối lượng tổng hợp tính từ chỉ số khối lượng cá thể:

$$I_q = \frac{\sum i_q (p_{i0} q_{i0})}{\sum p_{i0} q_{i0}} \cdot 100$$

Trong đó: i_q : chỉ số khối lượng cá thể mặt hàng i

$p_{i0}q_{i0}$: trọng số (doanh số bán mặt hàng i ở kỳ gốc)

Ví dụ: Có dữ liệu về tình hình tiêu thụ 4 mặt hàng như sau:

Mặt hàng	Đơn vị tính	Đơn giá 2010 (1000đ)	Lượng tiêu thụ 2010	Chỉ số lượng bán thể 2016/2010 i_p (%)
Cá thu	Kg	70	4000	150
Thịt bò	Kg	75	10000	120
Dầu ăn	Lít	10	2000	110
Mở heo	Kg	8	800	75

Chỉ số khối lượng tổng hợp:

$$I_q = \frac{\sum i_q (p_{i0} q_{i0})}{\sum p_{i0} q_{i0}} \cdot 100$$

$$= \frac{150(70 \cdot 4000) + 120(75 \cdot 10000) + \ldots + 75(8 \cdot 800)}{70 \cdot 4000 + 75 \cdot 10000 + \ldots + 8 \cdot 800} = 127{,}5\%$$

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Chương 9

PHÂN TÍCH PHƯƠNG SAI

Phân tích phương sai về thực chất là một phương pháp kiểm định tham số. Nó được dùng để kiểm định giả thuyết về sự giống nhau của nhiều tổng thể có cùng phân phối chuẩn hay gần chuẩn. Tuy nhiên, *phương pháp này tỏ ra hữu ích khi dùng để kiểm định sự giống nhau của nhiều tổng thể xét riêng trên một yếu tố nào đó trong khi cô lập ảnh hưởng của một số yếu tố khác lên cùng tổng thể đó.* Chẳng hạn, người ta muốn so sánh ảnh hưởng của yếu tố tổ chức sản xuất lên năng suất lao động trong khi cô lập ảnh hưởng của yếu tố giới tính. *Phương pháp phân tích phương sai được xây dựng dựa trên cơ sở phân tích phương sai chung thành tổng các phương sai nhân tố hợp thành.* Mức độ khác biệt giữa các phương sai nhân tố là cơ sở để bác bỏ hay chấp nhận giả thuyết.

9.1. Phân tích phương sai một yếu tố

Giả sử, xét theo một yếu tố nào đó (ví dụ như yếu tố thị trường, yếu tố loại quảng cáo, yếu tố loại sản phẩm…) có t tổng thể (t thị trường, t loại quảng cáo, t loại sản phẩm…) có cùng phân phối chuẩn, cùng phương sai chung (σ^2). Ta cần kiểm định giả thuyết sau:

Giả thuyết:

$H_o : \mu_1 = \mu_2 = \ldots = \mu_t$ (không có sự khác biệt theo yếu tố)

$H_1 : \exists\ i \neq j$ mà $\mu_i \neq \mu_j$ (có sự khác biệt theo yếu tố)

(với i, j =1, 2,…, t)

Để thực hiện kiểm định, trên mỗi tổng thể lấy một mẫu ngẫu nhiên độc lập với kích thước n_i (thường $n_i > 5$). Ta được bảng dữ liệu có dạng tổng quát sau:

Mẫu ngẫu nhiên từ tổng thể				
1	2	3	…	t
x_{11}	x_{21}	x_{31}	.	x_{t1}
x_{12}	x_{22}	x_{32}	.	x_{t2}
x_{13}	x_{23}	x_{33}	.	x_{t3}
...	...	...	...	...
x_{1n1}	x_{2n2}	x_{3n3}	.	x_{tnt}
$\overline{x}_1$	$\overline{x}_2$	$\overline{x}_3$	…	$\overline{x}_t$

Trong đó:

* x_{ij} : Quan sát thứ j của mẫu i
* $\overline{x}_i$: Số trung bình mẫu i
* n_i : Số quan sát của mẫu i, $n = \sum_{i=1}^{t} n_i$
* Số trung bình tất cả các mẫu: $\overline{x} = \dfrac{\sum_{i=1}^{t} \overline{x}_i n_i}{\sum_{i=1}^{t} n_i}$

Trên mẫu tổng quát ta tính được các tổng bình phương phản ảnh sự biến thiên của hiện tượng nghiên cứu như sau.

Tổng bình phương mọi sai khác trên các mẫu:

$$TSS = \sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}\right)^2$$

Tổng bình phương các sai khác giữa các mẫu:

$$SSB = \sum_{i=1}^{t}\left(\overline{X}_i - \overline{X}\right)^2 n_i$$

Tổng bình phương các sai khác nội bộ các mẫu:

$$SSW = \sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}_i\right)^2$$

Giả sử H_o đúng, $\mu_1 = \mu_2 = \ldots = \mu_t$. Các tổng thể nói trên có cùng phân phối chuẩn, cùng kỳ vọng toán, cùng phương sai nên hoàn toàn giống nhau. Các mẫu trên có thể xem như được lấy ngẫu nhiên từ một tổng thể chung có phân phối chuẩn với phương sai là σ^2. Theo phân phối Khi bình phương:

$$\frac{TSS}{\sigma^2} = \frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}\right)^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\frac{SSB}{\sigma^2} = \frac{\sum_{i=1}^{t}\left(\overline{X}_i - \overline{X}\right)^2 n_i}{\sigma^2} \sim \chi^2_{t-1}$$

$$\frac{SSW}{\sigma^2} = \sum_{i=1}^{t}\frac{\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}_i\right)^2}{\sigma^2} \sim \chi^2_{n-t} \qquad \left(do \sum_{i=1}^{t}\chi^2_{(n_i-1)} \sim \chi^2_{n-t}\right)$$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Theo tính chất phân phối Khi bình phương: $\chi^2_{t-1} + \chi^2_{n-t} = \chi^2_{n-1}$

Suy ra: $\frac{SSB}{\sigma^2} + \frac{SSW}{\sigma^2} = \frac{TSS}{\sigma^2}$

Hay: SSB + SSW = TSS

Theo phân phối Fisher-Snedecor:

$$F = \frac{\frac{\chi^2_{t-1}}{t-1}}{\frac{\chi^2_{n-t}}{n-t}} = \frac{\frac{SSB}{t-1}}{\frac{SSW}{n-t}} \sim F_{t-1,n-t}$$

Do đó, *tiêu chuẩn kiểm định:* $F = \frac{S^2_B}{S^2_W}$

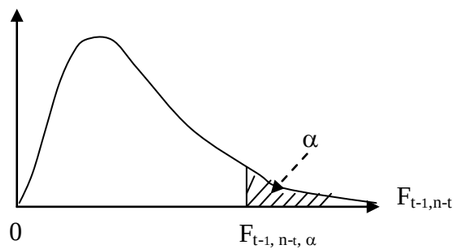
Trong đó: $S^2_B = \frac{SSB}{t-1}$, $S^2_W = \frac{SSW}{n-t}$

$$TSS = \sum_{i=1}^{t}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2 = n\left[\overline{x^2_{ij}} - (\bar{x})^2\right]$$

$$SSB = \sum_{i=1}^{t}(\bar{x}_i - \bar{x})^2 n_i = n\left[\overline{(\bar{x}_i)^2} - (\bar{x})^2\right]$$

SSW = TSS – SSB

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):



* $F \geq F_{t-1,n-t,\alpha}$ *: Bác bỏ H_o, chấp nhận H_1*

* $F < F_{t-1,n-t,\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

Ví dụ: Người ta muốn biết chi tiêu cho thực phẩm của khách du lịch ở 4 thị trường trọng điểm: Trung quốc, Nhật, Mỹ và Châu âu có khác nhau hay không. Để kiểm tra, người ta chọn ngẫu nhiên 8 khách đến từ mỗi thị trường. Số liệu về chi tiêu của họ thu được như sau:

Chi tiêu cho thực phẩm (nghìn USD)			
Trung quốc	Nhật	Mỹ	Châu âu
2	3	8	12
1	8	14	10
5	15	22	8
4	9	15	4
7	4	18	6
6	12	32	14
2	18	26	18
3	16	45	24
3,75	10,63	22,50	12,00

Kiểm định nhận định cho rằng chi tiêu trung bình của khách ở bốn thị trường trên là như nhau, với $\alpha = 0,05$.

Giả thuyết: $H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1 : \exists\, i \neq j$ mà $\mu_i \neq \mu_j$ (với i, j =1, 2, 3, 4)

Tiêu chuẩn kiểm định: $F = \frac{S^2_B}{S^2_W}$

Trong đó: $\bar{x} = \frac{\sum_{i=1}^{t}\bar{x}_i n_i}{\sum_{i=1}^{t} n_i} = \frac{\sum_{i=1}^{4}\bar{x}_i}{4} = 12,22$ (vì các n_i bằng nhau)

$$TSS = n\left[\overline{x^2_{ij}} - (\bar{x})^2\right]$$

$$= 32\left(\frac{2^2 + 1^2 + \ldots + 18^2 + 24^2}{32} - 12,22^2\right) = 2959,47$$

$$SSB = n\left[\overline{(\bar{x}_i)^2} - (\bar{x})^2\right]$$ (vì các n_i bằng nhau)

$$= 32\left(\frac{3,75^2 + 10,63^2 + 22,50^2 + 12,00^2}{4} - 12,22^2\right) = 1440,09$$

SSW = TSS - SSB = 2959,47 - 1440,09 = 1519,38

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

$$S_B^2 = \frac{SSB}{t-1} = \frac{1440,09}{4-1} = 480,03$$

$$S_W^2 = \frac{SSW}{n-t} = \frac{1519,09}{28} = 54,26$$

$$\Rightarrow \quad F = \frac{S_B^2}{S_W^2} = \frac{480,03}{54,26} = 8,85$$

Tra bảng: $F_{t-1,n-t,\alpha} = F_{3;28;0,05} = 2,95$

$F > F_{t-1,n-t,\alpha}$ => Bác bỏ H_o, chấp nhận H_1

Chú ý:

* Nếu kích thước các mẫu khác nhau, cần tính SSB có trọng số là các n_i.

$$SSB = n\left[\overline{(\bar{x}_i)^2} - (\bar{x})^2\right] = n\left[\frac{\sum_{i=1}^{t}(\bar{x}_i)^2 n_i}{\sum_{i=1}^{t} n_i} - (\bar{x})^2\right]$$

* Các giả định trong phương pháp phân tích phương sai là không quá nghiêm ngặt. Nghĩa là, nếu phân phối của các tổng thể là không quá lệch hay phương sai là không quá khác nhau thì các kết luận của phương pháp phân tích phương sai nói trên vẫn bảo đảm ý nghĩa trên các mẫu có kích thước bằng nhau và khá lớn.

* Để kiểm tra giả định phân phối chuẩn có thể dùng kiểm định Kolmogorov-smirnov trong SPSS.

* Để kiểm tra giả định phương sai bằng nhau, có thể sử dụng kiểm định Levene trong SPSS.

* Trường hợp dữ liệu không đáp ứng giả định về phân phối chuẩn hay giả định phương sai bằng nhau phải dùng phương pháp kiểm định hạng Kruskal-Wallis để thay thế.

9.2. Một số thiết kế thực nghiệm thường dùng trong phân tích phương sai

Trong thực hành nghiên cứu thống kê, *phương pháp phân tích phương sai có thể được vận dụng trong việc so sánh ảnh hưởng của các tác nhân khác nhau (treatments) của một yếu tố (factor) đến một tổng thể nghiên cứu*.

Ví dụ, người ta muốn so sánh tác dụng của ba phương cách quảng cáo đến tổng thể người tiêu dùng trên một thị trường. Yếu tố đang nghiên cứu là quảng cáo. Ba phương cách quảng cáo là ba tác nhân của nó. Tổng thể người tiêu dùng trên thị trường là tổng thể nghiên cứu.

Trong thực tế, *tác động đến tổng thể nghiên cứu thường không chỉ có một mình yếu tố đang nghiên cứu, đồng thời với nó, còn có thể có nhiều yếu tố khác nữa (gọi là yếu tố ngoại lai)*.

Ví dụ, khi quảng cáo, ngoài yếu tố quảng cáo còn có các yếu tố khác nữa đồng thời tác động đến người tiêu dùng như năng lực của (các) nhân viên quảng cáo, phong tục của người tiêu dùng (ở các địa phương khác nhau)...

Do đó, muốn so sánh chính xác các tác nhân của yếu tố đang nghiên cứu thì phải tìm cách loại trừ các ảnh hưởng khác nhau của các yếu tố khác. *Công việc thiết kế mẫu thực nghiệm nhằm loại trừ các ảnh hưởng khác nhau của các yếu tố khác, phục vụ phân tích phương sai so sánh các tác nhân của một yếu tố đang nghiên cứu nào đó, được gọi là thiết kế thực nghiệm.* Người ta thường dùng một số thiết kế thực nghiệm sau:

9.2.1. Thiết kế ngẫu nhiên hoàn toàn

Thiết kế ngẫu nhiên hoàn toàn được dùng để so sánh ảng hưởng giữa các tác nhân (xử lý) của một yếu tố nghiên cứu lên một hiện tượng mà không đề cập đến việc loại trừ một yếu tố ngoại lai cụ thể nào. Thiết kế ngẫu nhiên hoàn toàn được tiến hành như sau:

+ Từ tổng thể nghiên cứu chọn mẫu gốc ngẫu nhiên n_i đơn vị mẫu cho tác nhân i.

+ Cho mỗi tác nhân nghiên cứu tác động lên mẫu gốc tương ứng.

+ Đo lường kết quả trên từng đơn vị tổng thể của mẫu gốc sau tác động của mỗi tác nhân, ta thu được dữ liệu mẫu có dạng:

Mẫu của tác nhân				
1	2	3	…	t
x_{11}	x_{21}	x_{31}	.	x_{t1}
x_{12}	x_{22}	x_{32}	.	x_{t2}
x_{13}	x_{23}	x_{33}	.	x_{t3}
...	...	...	...	...
x_{1n1}	x_{2n2}	x_{3n3}	.	x_{tnt}

Giả thuyết về tác động giống nhau của các tác nhân *(xử lý)* đang nghiên cứu đến hiện tượng:

$H_o : \mu_1 = \mu_2 = \ldots = \mu_t$

$H_1 : \exists\, i \neq j$ mà $\mu_i \neq \mu_j$ (với i, j =1, 2,…, t)

Phương pháp phân tích phương sai một chiều (One-way analysis of variance) được *dùng cho thiết kế ngẫu nhiên hoàn toàn.*

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Chú ý: Thiết kế ngẫu nhiên hoàn toàn có thể được tạo ra từ một tổng thể thực tế đã chịu ảnh hưởng của yếu tố nghiên cứu. Từ tổng thể lấy một mẫu ngẫu nhiên kích thước n. Phân tổ mẫu này theo các tác nhân của yếu tố nghiên cứu.

9.2.2. Thiết kế khối ngẫu nhiên

Thiết kế khối ngẫu nhiên được dùng để so sánh ảng hưởng giữa các tác nhân của một yếu tố nghiên cứu lên một hiện tượng trong khi có nhắm đến loại trừ một yếu tố ngoại lai. Thiết kế khối ngẫu nhiên được tiến hành như sau:

+ Thiết kế này dựa trên một sơ đồ có t dòng, tương ứng với t tác nhân của yếu tố đang nghiên cứu *(xử lý)* và b cột, tương ứng với b tác nhân của yếu tố ngoại lai (khối).

Các tác nhân chính (xử lý)	Các tác nhân ngoại lai (khối)				
	B_1	B_2	B_3	…	B_b
T_1					
T_2					
T_3					
…					
T_t					

+ Từ tổng thể nghiên cứu chọn mẫu gốc ngẫu nhiên, một đơn vị mẫu cho một ô trong sơ đồ.

+ Cho mỗi cặp tác nhân chính-ngoại lai (T_i-B_j) tác động đồng thời lên đơn vị mẫu gốc tương ứng trong mỗi ô của sơ đồ trên.

+ Đo lường kết quả trên từng đơn vị tổng thể mẫu gốc ở mỗi ô sau tác động tương ứng của cặp tác nhân chính-ngoại lai ở ô đó, ta thu được dữ liệu mẫu có dạng:

Các tác nhân chính (xử lý)	Các tác nhân ngoại lai (khối)					Số trung bình
	B_1	B_2	B_3	…	B_b	
T_1	x_{11}	x_{12}	x_{13}	…	x_{1b}	$\overline{x}_{t_1}$
T_2	x_{21}	x_{22}	x_{23}	…	x_{2b}	$\overline{x}_{t_2}$
T_3	x_{31}	x_{32}	x_{33}	…	x_{3b}	$\overline{x}_{t_3}$
…	…	…	…	…	…	
T_t	x_{t1}	x_{t2}	x_{t3}	…	x_{tb}	$\overline{x}_{t_t}$
Số trung bình	$\overline{x}_{b_1}$	$\overline{x}_{b_2}$	$\overline{x}_{b_3}$		$\overline{x}_{b_b}$	$\overline{x}$

Trong thiết kế khối ngẫu nhiên nói trên, mỗi Số trung bình theo mỗi tác nhân chính (xử lý) chịu ảnh hưởng của tất cả các tác nhân ngoại lai (khối). Do đó, thiết kế này giúp chúng ta loại trừ ảnh hưởng khác nhau của các tác nhân ngoại lai (khối) không mong muốn đến các tác nhân chính (xử lý) khi so sánh chúng với nhau.

Phương pháp phân tích phương sai được dùng cho thiết kế khối ngẫu nhiên tiến hành như sau.

Giả thuyết (1) có sự giống nhau về ảnh hưởng của các xử lý đến hiện tượng nhiên cứu:

$H_o : \mu_{t1} = \mu_{t2} = \ldots = \mu_{tt}$

$H_1 : \exists\, i \neq j$ mà $\mu_{ti} \neq \mu_{tj}$ (với i, j =1, 2,…, t)

Giả thuyết (2) có sự giống nhau về ảnh hưởng của các khối đến hiện tượng nghiên cứu:

$H_o : \mu_{b1} = \mu_{b2} = \ldots = \mu_{bb}$

$H_1 : \exists\, i \neq j$ mà $\mu_{bi} \neq \mu_{bj}$ (với i, j =1, 2,…, b)

Phương pháp phân tích phương sai giả định tất cả các tổng thể có cùng phân phối chuẩn và cùng phương sai σ^2. Trên mẫu tổng quát:

Tổng bình phương mọi sai khác trên các mẫu:

$$TSS = \sum_{i=1}^{t}\sum_{j=1}^{b}(X_{ij} - \overline{X})^2$$

Tổng bình phương các sai khác giữa các xử lý:

$$SST = \sum_{i=1}^{t}(\overline{X}_{t_i} - \overline{X})^2 b$$

Tổng bình phương các sai khác giữa các khối:

$$SSB = \sum_{j=1}^{b}(\overline{X}_{b_j} - \overline{X})^2 t$$

Tổng bình phương các sai khác ngoài xử lý và khối (sai số):

$$SSE = TSS - SST - SSB$$

Chia hai vế cho σ^2, được:

$$\frac{SSE}{\sigma^2} = \frac{TSS}{\sigma^2} - \frac{SST}{\sigma^2} - \frac{SSB}{\sigma^2}$$

Nếu các giả thuyết (1) và (2) nói trên đúng, theo phân phối khi bình phương:

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

$$\frac{TSS}{\sigma^2} \sim \chi^2_{n-1} \quad , \quad \frac{SST}{\sigma^2} \sim \chi^2_{t-1} \quad , \quad \frac{SSB}{\sigma^2} \sim \chi^2_{b-1}$$

Do: $\chi^2_{n-1} - \chi^2_{t-1} - \chi^2_{b-1} = \chi^2_{(t-1)(b-1)} \quad => \quad \frac{SSE}{\sigma^2} \sim \chi^2_{(t-1)(b-1)}$

Theo phân phối Fisher-Snedecor:

$$F = \frac{\frac{\chi^2_{t-1}}{t-1}}{\frac{\chi^2_{(t-1)(b-1)}}{(t-1)(b-1)}} = \frac{\frac{SST}{t-1}}{\frac{SSE}{(t-1)(b-1)}} \sim F_{t-1,(t-1)(b-1)}$$

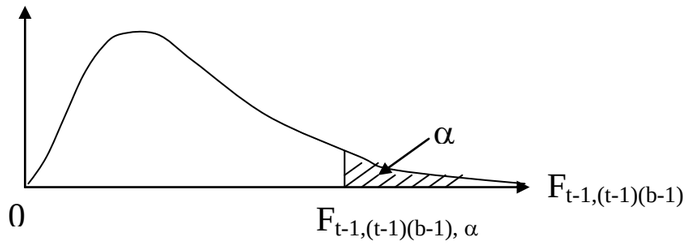
Do đó,*chọn tiêu chuẩn kiểm định giả thuyết (1):* $F = \frac{S^2_T}{S^2_E}$

Trong đó: $S^2_T = \frac{SST}{t-1}$, $S^2_E = \frac{SSE}{(t-1)(b-1)}$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):

* $F_T \geq F_{t-1,(t-1)(b-1),\alpha}$ *: Bác bỏ H_o, chấp nhận H_1*

* $F_T < F_{t-1,(t-1)(b-1),\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*



Tương tự, tiêu chuẩn kiểm định giả thuyết (2): $F_B = \frac{S^2_B}{S^2_E}$

Trong đó: $S^2_B = \frac{SSB}{b-1}$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):

* $F_B \geq F_{b-1,(t-1)(b-1),\alpha}$ *: Bác bỏ H_o, chấp nhận H_1*

* $F_B < F_{b-1,(t-1)(b-1),\alpha}$ *: Chưa đủ cơ sở bác bỏ Ho.*

Ví dụ: Một thực nghiệm được tiến hành để nghiên cứu ảnh hưởng của nghỉ ngơi giữa ca đến năng suất lao động trong một DN. Có bốn cách thức nghỉ ngơi là không nghỉ (T_1), nghỉ một lần (T_2), nghỉ hai lần (T_3) và nghỉ ba lần (T_4). Mỗi cách thức nghỉ ngơi được thử nghiệm trong năm ngày của một tuần. Người ta thấy thứ tự các ngày trong tuần có thể có ảnh hưởng khác nhau đến kết quả thực nghiệm. Vì vậy, một thiết kế khối ngẫu nhiên được thực hiện. Để tránh ảnh hưởng của các tuần thử nghiệm trước lên kết quả của tuần thử nghiệm sau, các tuần thử nghiệm được chọn cách nhau ba tuần. Kết quả thu được về năng suất trong năm ngày thử nghiệm (SP/giờ) cho mỗi cách thức nghỉ ngơi như sau:

Cách thức nghỉ ngơi (nhóm)	Ngày trong tuần (khối)					Số trung bình
	Thứ 2 (B_1)	Thứ 3 (B_2)	Thứ 4 (B_3)	Thứ 5 (B_4)	Thứ 6 (B_5)	
T_1	16	15	15	13	12	14,2
T_2	18	17	17	16	14	16,4
T_3	17	18	15	16	15	16,2
T_4	14	14	13	13	14	13,6
Số trung bình	16,25	16,00	15,00	14,50	13,75	15,1

Kiểm định giả thuyết về ảnh hưởng giống nhau của các cánh thức nghỉ ngơi đến năng suất lao động với $\alpha = 0,05$.

$H_o : \mu_{t1} = \mu_{t2} = \mu_{t3} = \mu_{t4}$

$H_1 : \exists\, i \neq j$ mà $\mu_{ti} \neq \mu_{tj}$ (với i, j =1, 2, 3, 4)

Tiêu chuẩn kiểm định: $F_T = \frac{S^2_T}{S^2_E}$

Trong đó:

$$TSS = n\left[\overline{x^2_{ij}} - (\overline{x})^2\right]$$

$$= 20\left(\frac{16^2 + 18^2 + \ldots + 15^2 + 14^2}{20} - 15,1^2\right) = 57,8$$

$$SST = n\left[\overline{(\overline{x}_{t_i})^2} - (\overline{x})^2\right]$$

$$= 20\left(\frac{14,2^2 + 16,4^2 + 16,2^2 + 13,6^2}{4} - 15,1^2\right) = 29,8$$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

$$SSB = n\left[\overline{(\bar{x}_{b_j})^2} - (\bar{x})^2\right]$$

$$= 20\left(\frac{16{,}15^2 + 16^2 + 15^2 + 14{,}5^2 + 13{,}75^2}{5} - 15{,}1^2\right) = 17{,}3$$

SSE = TSS - SST - SSB = 57,8 - 29,8 - 17,3 = 10,7

$$S_T^2 = \frac{SST}{t-1} = \frac{57{,}8}{4-1} = 9{,}93$$

$$S_E^2 = \frac{SSE}{(t-1)(b-1)} = \frac{10{,}7}{(4-1)(5-1)} = 0{,}89$$

$$\Rightarrow \quad F_T = \frac{S_T^2}{S_E^2} = \frac{9{,}93}{0{,}89} = 11{,}14$$

Tra bảng: $F_{t-1,(t-1)(b-1),\alpha} = F_{3;12;0,05} = 3{,}49$

$F_T > F_{t-1,(t-1)(b-1),\alpha}$ => Bác bỏ H_o, chấp nhận H_1

Tương tự, kiểm định giả thuyết có sự giống nhau về ảnh hưởng của thứ tự các ngày trong tuần đến năng suất:

$H_o : \mu_{b1} = \mu_{b2} = \mu_{b3} = \mu_{b24} = \mu_{b5}$

$H_1 : \exists\, i \neq j$ với i, j =1, 2,…, 5 mà $\mu_{bi} \neq \mu_{bj}$

Tiêu chuẩn kiểm định: $F_B = \frac{S_B^2}{S_E^2}$

Trong đó: $S_B^2 = \frac{SSB}{b-1} = \frac{17{,}3}{5-1} = 4{,}32$

$$\Rightarrow \quad F_B = \frac{S_B^2}{S_E^2} = \frac{4{,}32}{0{,}89} = 4{,}85$$

Tra bảng: $F_{b-1,(t-1)(b-1),\alpha} = F_{4;12;0,05} = 3{,}26$

$F_B > F_{b-1,(t-1)(b-1),\alpha}$ => Bác bỏ H_o , chấp nhận H_1

Chú ý:

* Thiết kế khối ngẫu nhiên cũng có thể được tạo ra từ một tổng thể thực tế đã chịu ảnh hưởng của yếu tố nghiên cứu và yếu tố ngoại lai. Từ tổng thể lấy một mẫu ngẫu nhiên kích thước m>n. Phân tổ kết hợp mẫu này theo cả hai yếu tố, thành t tổ xử lý và b tổ khối. Như vậy có t*b tổ kết hợp tương ứng với t*b ô như sơ đồ thiết kế khối nói trên. Trên mỗi tổ kết hợp (ô) lấy ngẫu nhiên một đơn vị mẫu để đưa vào sơ đồ thiết kế khối tương ứng nói trên.

* Phương pháp phân tích phương sai dùng cho thiết kế khối ngẫu nhiên thường được gọi là phân tích phương sai hai chiều không lặp (Two-way analysis of variance without replication).

9.2.3. Thiết kế vuông ngẫu nhiên

Thiết kế vuông ngẫu nhiên được dùng để so sánh ảng hưởng giữa các tác nhân (xử lý) của một yếu tố nghiên cứu lên một hiện tượng trong khi có nhắm đến loại trừ hai yếu tố ngoại lai. Thiết kế vuông ngẫu nhiên được tiến hành như sau:

+ Thiết kế này dựa trên một sơ đồ có t dòng, tương ứng với t tác nhân của yếu tố ngoại lai thứ nhất và t cột, tương ứng với t tác nhân của yếu tố ngoại lai thứ hai. Giao của chúng là (t* t) ô. t tác nhân của yếu tố chính đang nghiên cứu được bố trí vào các ô sao cho có một và chỉ một tác nhân chính trên một dòng hay một cột. Để minh hoạ, một sơ đồ thiết kế vuông của bốn tác nhân chính là T_1, T_2, T_3, T_4 như sau:

Các tác nhân ngoại lai thứ nhất (dòng)	Các tác nhân ngoại lai thứ hai (cột)			
	C_1	C_2	C_3	C_4
R_1	T_1	T_2	T_3	T_4
R_2	T_2	T_3	T_4	T_1
R_3	T_3	T_4	T_1	T_2
R_4	T_4	T_1	T_2	T_3

+ Từ tổng thể nghiên cứu chọn mẫu gốc ngẫu nhiên, một đơn vị mẫu cho một ô của sơ đồ trên.

+ Cho mỗi bộ tác nhân "chính-ngoại lai1-ngoại lai2" tác động đồng thời lên đơn vị tổng thể mẫu gốc tương ứng với mỗi ô trong sơ đồ trên.

+ Đo lường kết quả trên từng đơn vị tổng thể mẫu gốc ở mỗi ô sau tác động tương ứng của bộ tác nhân "chính-ngoại lai1-ngoại lai2" ở ô đó, ta thu được dữ liệu mẫu có dạng:

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Các tác nhân ngoại lai thứ nhất (Dòng)	Các tác nhân ngoại lai thứ hai (Cột)				Số trung bình
	C_1	C_2	C_3	C_4	
R_1	$x_{11(1)}$	$x_{12\,(2)}$	$x_{13\,(3)}$	$x_{14\,(4)}$	$\overline{x}_{r_1}$
R_2	$x_{21(2)}$	$x_{22\,(3)}$	$x_{23\,(4)}$	$x_{24\,(1)}$	$\overline{x}_{r_2}$
R_3	$x_{31(3)}$	$x_{32\,(4)}$	$x_{33\,(1)}$	$x_{34\,(2)}$	$\overline{x}_{r_3}$
R_4	$x_{41(4)}$	$x_{42\,(1)}$	$x_{43\,(2)}$	$x_{44\,(3)}$	$\overline{x}_{r_4}$
Số trung bình	$\overline{x}_{c_1}$	$\overline{x}_{c_2}$	$\overline{x}_{c_3}$	$\overline{x}_{c_4}$	$\overline{x}$

(số trong ngoặc chỉ xử lý)

Các Số trung bình của các tác nhân chính (xử lý) là:

$$\overline{x}_{t_1} = \frac{x_{11(1)} + x_{42(1)} + x_{33(1)} + x_{24(1)}}{4}$$

$$\overline{x}_{t_2} = \frac{x_{21(2)} + x_{12(2)} + x_{43(2)} + x_{34(2)}}{4}$$

$$\overline{x}_{t_3} = \frac{x_{31(3)} + x_{22(3)} + x_{13(3)} + x_{44(3)}}{4}$$

$$\overline{x}_{t_4} = \frac{x_{41(4)} + x_{32(4)} + x_{23(4)} + x_{14(4)}}{4}$$

Trong thiết kế vuông ngẫu nhiên nói trên, mỗi Số trung bình theo mỗi tác nhân chính (xử lý) chịu ảnh hưởng của tất cả các tác nhân của yếu tố ngoại lai thứ nhất (dòng) và yếu tố ngoại lai thứ hai (cột). Do đó, thiết kế này giúp chúng ta loại trừ ảnh hưởng khác nhau của các tác nhân của hai yếu tố ngoại lai không mong muốn đến các tác nhân chính (xử lý) khi so sánh chúng với nhau.

Phương pháp phân tích phương sai dùng cho thiết kế vuông ngẫu nhiên tiến hành như sau.

Giả thuyết (1) về tác động giống nhau của các tác nhân chính (xử lý):

$H_o : \mu_{t1} = \mu_{t2} = \ldots = \mu_{tt}$

$H_1 : \exists\ i \neq j$ mà $\mu_{ti} \neq \mu_{tj}$ (với i, j =1, 2,…, t)

Giả thuyết (2) về tác động giống nhau của các tác nhân ngoại lai thứ nhất:

$H_o : \mu_{r1} = \mu_{r2} = \ldots = \mu_{rt}$

$H_1 : \exists\ i \neq j$ mà $\mu_{ri} \neq \mu_{rj}$ (với i, j =1, 2,…, t)

Giả thuyết (3) về tác động giống nhau của các tác nhân ngoại lai thứ hai:

$H_o : \mu_{c1} = \mu_{c2} = \ldots = \mu_{ct}$

$H_1 : \exists\ i \neq j$ mà $\mu_{ci} \neq \mu_{cj}$ (với i, j =1, 2,…, t)

Phương pháp phân tích phương sai giả định tất cả các tổng thể có cùng phân phối chuẩn và cùng phương sai σ^2. Trên mẫu tổng quát:

Tổng bình phương mọi sai khác trên các mẫu:

$$TSS = \sum_{i=1}^{t}\sum_{j=1}^{t}(X_{ij} - \overline{X})^2$$

Tổng bình phương các sai khác giữa các xử lý:

$$SST = \sum_{k=1}^{t}(\overline{X}_{t_k} - \overline{X})^2 t$$

Tổng bình phương các sai khác giữa các dòng:

$$SSR = \sum_{i=1}^{t}(\overline{X}_{r_i} - \overline{X})^2 t$$

Tổng bình phương các sai khác giữa các cột:

$$SSC = \sum_{j=1}^{t}(\overline{X}_{c_j} - \overline{X})^2 t$$

Tổng bình phương các sai khác ngoài xử lý, dòng và cột (sai số):

$$SSE = TSS - SST - SSR - SSC$$

Chia hai vế cho σ^2, được:

$$\frac{SSE}{\sigma^2} = \frac{TSS}{\sigma^2} - \frac{SST}{\sigma^2} - \frac{SSR}{\sigma^2} - \frac{SSC}{\sigma^2}$$

Nếu các giả thuyết (1), (2) và (3) nói trên đúng, theo phân phối khi bình phương:

$$\frac{TSS}{\sigma^2} \sim \chi^2_{n-1}\ ,\quad \frac{SST}{\sigma^2} \sim \chi^2_{t-1}\ ,\quad \frac{SSR}{\sigma^2} \sim \chi^2_{t-1}\ ,\quad \frac{SSC}{\sigma^2} \sim \chi^2_{t-1}$$

Do: $\chi^2_{n-1} - \chi^2_{t-1} - \chi^2_{t-1} - \chi^2_{t-1} = \chi^2_{(t-1)(t-2)}$ => $\frac{SSE}{\sigma^2} \sim \chi^2_{(t-1)(t-2)}$

Theo phân phối Fisher-Snedecor:

$$F = \frac{\frac{\chi^2_{t-1}}{t-1}}{\frac{\chi^2_{(t-1)(t-2)}}{(t-1)(t-2)}} = \frac{\frac{SST}{t-1}}{\frac{SSE}{(t-1)(t-2)}} \sim F_{t-1,(t-1)(t-2)}$$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

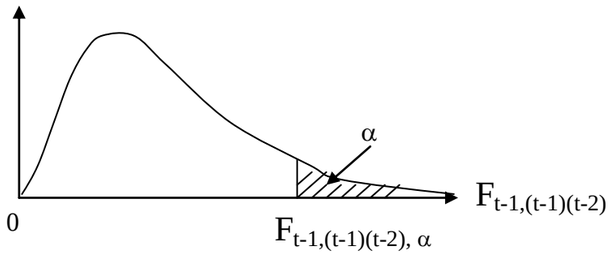
Do đó,*chọn tiêu chuẩn kiểm định giả thuyết (1):* $F_T = \frac{S_T^2}{S_E^2}$

Trong đó: $S_T^2 = \frac{SST}{t-1}$, $S_E^2 = \frac{SSE}{(t-1)(t-2)}$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):

* $F_T \geq F_{t-1,(t-1)(t-2),\alpha}$ *: Bác bỏ H_o, chấp nhận H_1*

* $F_T < F_{t-1,(t-1)(t-2),\alpha}$ *: Chưa đủ cơ sở bác bỏ H_o.Tạm chấp nhận H_o*



Tương tự, *chọn tiêu chuẩn kiểm định giả thuyết (2):* $F_R = \frac{S_R^2}{S_E^2}$

Trong đó: $S_R^2 = \frac{SSR}{t-1}$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):

* $F_R \geq F_{t-1,(t-1)(t-2),\alpha}$ *: Bác bỏ H_o, chấp nhận H_1*

* $F_R < F_{t-1,(t-1)(t-2),\alpha}$ *: Chưa đủ cơ sở bác bỏ H_o.Tạm chấp nhận H_o*

Tương tự, *tiêu chuẩn kiểm định giả thuyết (3):* $F_C = \frac{S_C^2}{S_E^2}$

Trong đó: $S_C^2 = \frac{SSC}{t-1}$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):

* $F_C \geq F_{t-1,(t-1)(t-2),\alpha}$ *: Bác bỏ H_o, chấp nhận H_1*

* $F_C < F_{t-1,(t-1)(t-2),\alpha}$ *: Chưa đủ cơ sở bác bỏ H_o.Tạm chấp nhận H_o*

Ví dụ: Một nhà kinh doanh thương mại muốn so sánh hiệu quả của bốn cách thức bày hàng: trên kệ, ở ngang tầm mắt (T_1), cuối kệ (T_2), trên các tầng giữa các kệ (T_3) và, treo trên dây (T_4). Các cửa hàng khác nhau và thứ tự tuần thử nghiệm có thể có ảnh hưởng khác nhau đến kết quả thử nghiệm. Do đó một thiết kế vuông được sử dụng. Doanh số bán trong các tuần thử nghiệm (triệu đồng) ở mỗi cửa hàng thu được như sau:

Cửa hàng (Row)	Thứ tự tuần thử nghiệm (Column)				Số trung bình
	1	2	3	4	
1	(T_1) 35,5	(T_2) 42,5	(T_3) 52,5	(T_4) 34,0	41,125
2	(T_2) 40,5	(T_3) 48,5	(T_4) 33,0	(T_1) 38,0	40,0
3	(T_3) 46,0	(T_4) 34,5	(T_1) 36,5	(T_2) 45,5	40,625
4	(T_4) 32,0	(T_1) 34,0	(T_2) 41,0	(T_3) 56,0	40,75
Số trung bình	38,5	39,875	40,75	43,375	40,625

Kiểm định giả thuyết về tác dụng giống nhau của các cách thức bày hàng, với $\alpha = 0,05$.

$H_o : \mu_{t1} = \mu_{t2} = \mu_{t3} = \mu_{t4}$

$H_1 : \exists\, i \neq j$ mà $\mu_{ti} \neq \mu_{tj}$ (với i, j =1, 2,…, 4)

$$\bar{x}_{t_1} = \frac{35,5 + 34,0 + 36,5 + 38,0}{4} = 36$$

$$\bar{x}_{t_2} = \frac{40,5 + 42,5 + 41,0 + 45,5}{4} = 42,375$$

$$\bar{x}_{t_3} = \frac{46 + 48,5 + 52,5 + 56}{4} = 50,75$$

$$\bar{x}_{t_4} = \frac{32 + 34,5 + 33 + 34}{4} = 33,375$$

Tiêu chuẩn kiểm định: $F_T = \frac{S_T^2}{S_E^2}$

Trong đó:

$$TSS = n\left[\overline{x_{ij}^2} - (\bar{x})^2\right]$$

$$= 16\left(\frac{35,5^2 + 40,5^2 + \ldots + 45,5^2 + 56^2}{16} - 40,625^2\right) = 803,75$$

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

$$SST = n\left[\overline{(\overline{x}_{t_k})^2} - (\overline{x})^2\right]$$

$$= 16\left(\frac{36^2 + 42{,}375^2 + 50{,}75^2 + 33{,}375^2}{4} - 40{,}625^2\right) = 718{,}125$$

$$SSR = n\left[\overline{(\overline{x}_{r_i})^2} - (\overline{x})^2\right]$$

$$= 16\left(\frac{41{,}125^2 + 40^2 + 40{,}625^2 + 40{,}75^2}{4} - 40{,}625^2\right) = 2{,}625$$

$$SSC = n\left[\overline{(\overline{x}_{c_j})^2} - (\overline{x})^2\right]$$

$$= 16\left(\frac{38{,}5^2 + 39{,}875^2 + 40{,}75^2 + 43{,}375^2}{4} - 40{,}625^2\right) = 50{,}625$$

SSE = TSS - SST - SSR - SSC =803,75 - 718,125 - 2,625 - 50,625
= 32,375

$$S_T^2 = \frac{SST}{t-1} = \frac{718{,}125}{4-1} = 239{,}375$$

$$S_E^2 = \frac{SSE}{(t-1)(t-2)} = \frac{32{,}375}{(4-1)(4-2)} = 5{,}396$$

$$F_T = \frac{S_T^2}{S_E^2} = \frac{239{,}375}{5{,}396} = 44{,}36$$

Tra bảng: $F_{t-1,(t-1)(t-2),\alpha} = F_{3;6;0,05} = 4{,}76$

$F_T > F_{t-1,(t-1)(t-2),\alpha}$: Bác bỏ H_o, chấp nhận H_1

Tương tự, giả thuyết có sự giống nhau về tác động của các cửa hàng đến doanh số:

$H_o : \mu_{r1} = \mu_{r2} = \mu_{r3} = \mu_{r4}$

$H_1 : \exists\, i \neq j$ mà $\mu_{ri} \neq \mu_{rj}$ (với i, j =1, 2,…, 4)

Tiêu chuẩn kiểm định: $F_R = \frac{S_R^2}{S_E^2}$

Trong đó: $S_R^2 = \frac{SSR}{t-1} = \frac{2{,}625}{4-1} = 0{,}875$

$$F_R = \frac{S_R^2}{S_E^2} = \frac{0{,}875}{5{,}396} = 0{,}162$$

Tra bảng: $F_{t-1,(t-1)(t-2),\alpha} = F_{3;6;0,05} = 4{,}76$

$F_R < F_{t-1,(t-1)(t-2),\alpha}$ => Chưa đủ cơ sở bác bỏ H_o.Tạm chấp nhận H_o

Tương tự, ta cũng có thể kiểm định giả thuyết về tác động giống nhau của thứ tự các tuần thử nghiệm đến doanh số:

$H_o : \mu_{c1} = \mu_{c2} = \mu_{c3} = \mu_{c4}$

$H_1 : \exists\, i \neq j$ mà $\mu_{ci} \neq \mu_{cj}$ (với i, j =1, 2,…, 4)

Tiêu chuẩn kiểm định: $F_C = \frac{S_C^2}{S_E^2}$

Trong đó: $S_C^2 = \frac{SSC}{t-1} = \frac{50{,}625}{4-1} = 16{,}875$

$$F_C = \frac{S_C^2}{S_E^2} = \frac{16{,}875}{5{,}396} = 3{,}127$$

Tra bảng: $F_{t-1,(t-1)(t-2),\alpha} = F_{3;6;0,05} = 4{,}76$

$F_C < F_{t-1,(t-1)(t-2),\alpha}$: Chưa đủ cơ sở bác bỏ H_o.Tạm chấp nhận H_o

Chú ý:

* Thiết kế vuông ngẫu nhiên cũng có thể được tạo ra từ một tổng thể thực tế đã chịu ảnh hưởng của yếu tố nghiên cứu và hai yếu tố ngoại lai như sau. Từ tổng thể lấy một mẫu ngẫu nhiên kích thước m>n. Phân tổ kết hợp mẫu chung này theo cả ba yếu tố thành các tổ theo dòng, cột và ô như một sơ đồ thiết kế vuông nói trên. Trên mỗi tổ kết hợp (ô) lấy ngẫu nhiên một đơn vị mẫu để đưa vào một ô của sơ đồ thiết kế vuông tương ứng nói trên.

* Phương pháp phân tích phương sai dùng cho thiết kế vuông ngẫu nhiên thường được gọi là phân tích phương sai ba chiều không lặp (Three-way analysis of variance without replication).

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

9.2.4. Thiết kế thừa số ngẫu nhiên

Thiết kế thừa số ngẫu nhiên được dùng để so sánh ảng hưởng giữa các tác nhân (xử lý) của một yếu tố lên một hiện tượng, trong đó có nhắm đến loại trừ ảnh hưởng các yếu tố nghiên cứu khác; đồng thời nghiên cứu tương tác giữa các tác nhân của các yếu tố này.

Thiết kế thừa số ngẫu nhiên đơn giản nhất gồm có hai yếu tố được tiến hành như sau:

+ Thiết kế này dựa trên một sơ đồ có a dòng, tương ứng với a tác nhân của yếu tố thứ nhất và b cột, tương ứng với b tác nhân của yếu tố thứ hai, ô giao của dòng i và cột j tương ứng với tác nhân cặp (A_iB_j);

+ Từ tổng thể nghiên cứu chọn mẫu gốc ngẫu nhiên kích thước $n=a*b*c$, với c đơn vị tổng thể cho một ô của sơ đồ trên;

+ Cho cặp tác nhân (A_iB_j) tác động lên từng đơn vị tổng thể mẫu gốc trong ô (i,j) trong sơ đồ trên.

Các tác nhân của yếu tố thứ nhất (A)	Các tác nhân của yếu tố thứ hai (B)				
	B_1	B_2	B_3	…	B_b
A_1					
A_2					
A_3					
…					
A_a					

+ Đo lường kết quả, ta thu được dữ liệu mẫu có dạng:

Các tác nhân của yếu tố thứ nhất (A)	Các tác nhân của yếu tố thứ hai (B)			
	B_1	B_2	…	B_b
A_1	$x_{11(1)}$ $x_{11(2)}$ … $x_{11(c)}$	$x_{12(1)}$ $x_{12(2)}$ … $x_{12(c)}$	…	$x_{1b(1)}$ $x_{1b(2)}$ … $x_{1b(c)}$
A_2	$x_{21(1)}$ $x_{21(2)}$ … $x_{21(c)}$	$x_{22(1)}$ $x_{22(2)}$ … $x_{22(c)}$	…	$x_{2b(1)}$ $x_{2b(2)}$ … $x_{2b(c)}$
…	…	…	…	…
A_a	$x_{a1(1)}$ $x_{a1(2)}$ … $x_{a1(c)}$	$x_{a2(1)}$ $x_{a2(2)}$ … $x_{a2(c)}$	…	$x_{ab(1)}$ $x_{ab(2)}$ … $x_{ab(c)}$

(số trong ngoặc chỉ thứ tự đơn vị tổng thể trong mỗi ô)

Từ bảng trên, ta lập bảng dữ liệu các số trung bình ô, số trung bình dòng, trung bình cột:

Các tác nhân của yếu tố thứ nhất (A)	Các tác nhân của yếu tố thứ hai (B)				Số trung bình
	B_1	B_2	…	B_b	
A_1	$\overline{x}_{11}$	$\overline{x}_{12}$	…	$\overline{x}_{1b}$	$\overline{x}_{a_1}$
A_2	$\overline{x}_{21}$	$\overline{x}_{22}$	…	$\overline{x}_{2b}$	$\overline{x}_{a_2}$
…	…	…	…	…	…
A_a	$\overline{x}_{a1}$	$\overline{x}_{a2}$	…	$\overline{x}_{ab}$	$\overline{x}_{a_a}$
Số trung bình	$\overline{x}_{b_1}$	$\overline{x}_{b_2}$	…	$\overline{x}_{b_b}$	$\overline{x}$

Trong thiết kế nói trên, mỗi Số trung bình theo mỗi tác nhân chính $\overline{x}_{a_i}$ chịu ảnh hưởng của tất cả các tác nhân B_j. Và, mỗi Số trung bình theo mỗi tác nhân chính $\overline{x}_{b_j}$ chịu ảnh hưởng của tất cả các tác nhân A_i. Do đó, thiết kế này giúp chúng ta so sánh ảnh hưởng riêng của các tác nhân của mỗi yếu tố đến tổng thể nghiên cứu.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Phương pháp phân tích phương sai cho thiết kế thừa số ngẫu nhiên hai chiều được tiến hành như sau:

Giả thuyết (1): Các cách thức A_i tác động giống nhau đến hiện tượng nghiên cứu:

$$H_o : \mu_{a1} = \mu_{a2} = \ldots = \mu_{aa}$$

$$H_1 : \exists\, j \neq i \text{ mà } \mu_{ai} \neq \mu_{aj} \qquad (\text{với } j, i = 1, 2, \ldots, a)$$

Giả thuyết (2): Các cách thức B_j tác động giống nhau đến hiện tượng nghiên cứu:

$$H_o : \mu_{b1} = \mu_{b2} = \ldots = \mu_{bb}$$

$$H_1 : \exists\, i \neq j \text{ mà } \mu_{bj} \neq \mu_{bi} \qquad (\text{với } i, j = 1, 2, \ldots, b)$$

Giả thuyết (3): Không tồn tại tương tác giữa yếu tố A và yếu tố B.

Phương pháp phân tích phương sai giả định tất cả các tổng thể có cùng phân phối chuẩn và cùng phương sai σ^2. Trên mẫu tổng quát:

Tổng bình phương mọi sai khác trên các mẫu:

$$TSS = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(X_{ijk} - \overline{X})^2$$

Tổng bình phương các sai khác do ảnh hưởng riêng của yếu tố A:

$$SSA = \sum_{i=1}^{a}(\overline{X}_{a_i} - \overline{X})^2 bc$$

Tổng bình phương các sai khác do ảnh hưởng riêng của yếu tố B:

$$SSB = \sum_{j=1}^{b}(\overline{X}_{b_j} - \overline{X})^2 ac$$

Tổng bình phương các sai khác do ảnh hưởng kết hợp yếu tố A và B:

$$SS(AB) = \sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{X}_{ij} - \overline{X})^2 c$$

Tổng bình phương các sai khác do tương tác của yếu tố A và yếu tố B:

$$SSAB = SS(AB) - SSA - SSB$$

Tổng bình phương các sai khác còn lại (sai số):

$$SSE = TSS - SS(AB)$$

Chia hai vế hai biểu thức sau cùng cho σ^2, được:

$$\frac{SSAB}{\sigma^2} = \frac{SS(AB)}{\sigma^2} - \frac{SSA}{\sigma^2} - \frac{SSB}{\sigma^2}$$

Và:
$$\frac{SSE}{\sigma^2} = \frac{TSS}{\sigma^2} - \frac{SS(AB)}{\sigma^2}$$

Nếu các giả thuyết (1), (2) và (3) nói trên đúng, theo phân phối khi bình phương:

$$\frac{TSS}{\sigma^2} \sim \chi^2_{n-1} \quad , \quad \frac{SSA}{\sigma^2} \sim \chi^2_{a-1} \quad , \quad \frac{SSB}{\sigma^2} \sim \chi^2_{b-1} \quad , \quad \frac{SS(AB)}{\sigma^2} \sim \chi^2_{ab-1}$$

Do: $\chi^2_{ab-1} - \chi^2_{a-1} - \chi^2_{b-1} = \chi^2_{(a-1)(b-1)}$ => $\frac{SSAB}{\sigma^2} \sim \chi^2_{(a-1)(b-1)}$

Và: $\chi^2_{n-1} - \chi^2_{ab-1} = \chi^2_{n-ab}$ => $\frac{SSE}{\sigma^2} \sim \chi^2_{n-ab}$

Theo phân phối Fisher-Snedecor:

$$F = \frac{\frac{\chi^2_{a-1}}{a-1}}{\frac{\chi^2_{n-ab}}{n-ab}} = \frac{\frac{SSA}{a-1}}{\frac{SSE}{n-ab}} \sim F_{a-1,\,n-ab}$$

$$F = \frac{\frac{\chi^2_{b-1}}{b-1}}{\frac{\chi^2_{n-ab}}{n-ab}} = \frac{\frac{SSB}{b-1}}{\frac{SSE}{n-ab}} \sim F_{b-1,\,n-ab}$$

$$F = \frac{\frac{\chi^2_{(a-1)(b-1)}}{(a-1)(b-1)}}{\frac{\chi^2_{n-ab}}{n-ab}} = \frac{\frac{SSAB}{(a-1)(b-1)}}{\frac{SSE}{n-ab}} \sim F_{(a-1)(b-1),\,n-ab}$$

Do đó, chọn t*iêu chuẩn kiểm định giả thuyết (1):* $F_A = \frac{S^2_A}{S^2_E}$

Trong đó:

$$S^2_A = \frac{SSA}{a-1} \quad , \quad S^2_B = \frac{SSB}{b-1}$$

$$S^2_{AB} = \frac{SSAB}{(a-1)(b-1)} \quad , \quad S^2_E = \frac{SSE}{n-ab}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

* $F_A \geq F_{(a-1),n-ab,\alpha}$ *: Bác bỏ* H_o*, chấp nhận* H_1

* $F_A < F_{(a-1),n-ab,\alpha}$ *: Chưa đủ cơ sở bác bỏ* H_o*.Tạm chấp nhận* H_o

Tiêu chuẩn kiểm định giả thuyết (2):

$$F_B = \frac{S_B^2}{S_E^2}$$

Ta kết luận với mức ý nghĩa α *(xác suất sai lầm loại I):*

* $F_B \geq F_{(b-1),n-ab,\alpha}$ *: Bác bỏ Ho, chấp nhận* H_1

* $F_B < F_{(b-1),n-ab,\alpha}$ *: Chưa đủ cơ sở bác bỏ* H_o*.Tạm chấp nhận* H_o

Tiêu chuẩn kiểm định giả thuyết (3): $F_{AB} = \frac{S_{AB}^2}{S_E^2}$

Ta kết luận với mức ý nghĩa α *(xác suất sai lầm loại I):*

* $F_{AB} \geq F_{(a-1)(b-1),n-ab,\alpha}$ *: Giữa A và B tồn tại sự tương tác cặp*

* $F_{AB} < F_{(a-1)(b-1),n-ab,\alpha}$ *: Giữa A và B không tồn tại tương tác cặp*

Chú ý:

** Tương tác giữa hai yếu tố A, B là phần tác động mạnh hơn hay yếu hơn so với tổng tác động riêng của yếu tố A và yếu tố B lên hiện tượng nghiên cứu.*

* Thiết kế thừa số ngẫu nhiên hai chiều cũng có thể được tạo ra từ một tổng thể thực tế đã chịu ảnh hưởng của hai yếu tố nghiên cứu như sau. Từ tổng thể lấy một mẫu ngẫu nhiên kích thước m>n. Phân tổ kết hợp mẫu chung này theo cả hai yếu tố thành a tổ của yếu tố thứ nhất và b tổ của yếu tố thứ hai.

* Phương pháp phân tích phương sai dùng cho thiết kế thừa số ngẫu nhiên hai chiều thường được gọi là phân tích phương sai hai chiều lặp (Two-way analysis of variance with replication).

Ví dụ: Một doanh nghiệp thương mại muốn nghiên cứu tác động đồng thời của các chính sánh giá và các cách thức bày hàng đến doanh số. Có ba chính sách giá là: giá thường (A_1), giá giảm (A_2) và giá có khuyến mại (A_3). Có ba cách thức bày hàng là: không gian thường (B_1), không gian thường có chặn lối đi ở giữa (B_2) và không gian gấp đôi không gian thường (B_3). Do đó, một thiết kế thừa số đã được sử dụng. Mỗi kết hợp (A_iB_j) được thử nghiệm bán hàng trong ba tuần. Doanh số (chục triệu đồng) thu được như sau:

Cách thức bày hàng	Chính sách giá		
	A_1	A_2	A_3
B_1	9,5	12,5	16,0
	10,8	13,5	17,0
	11,0	13,0	17,5
B_2	12,0	18,5	25,5
	13,5	20,0	27,0
	12,5	21,0	26,5
B_3	13,0	16,0	19,0
	12,5	15,0	21,5
	12,0	14,5	20,0

Kiểm định các giả thuyết với $\alpha = 0,05$.

Ta có bảng dữ liệu các Số trung bình:

Cách thức bày hàng	Chính sáchgiá			Số trung bình
	A_1	A_2	A_3	
B_1	10,43	13,00	16,83	13,42
B_2	12,67	19,83	26,33	19,61
B_3	12,50	15,17	20,17	15,94
Số trung bình	11,87	16,00	21,11	16,32

$$TSS = n\left[\overline{x_{ijk}^2} - (\overline{x})^2\right]$$

$$= 27\left(\frac{9,5^2 + 10,8^2 + ... + 21,5^2 + 20^2}{27} - 16,32^2\right) = 621,17$$

$$SS(AB) = n\left[\overline{(\overline{x}_{ij})^2} - (\overline{x})^2\right]$$

$$= 27\left(\frac{10,43^2 + 12,67^2 + ... + 26,33^2 + 20,17^2}{9} - 16,32^2\right) = 607,84$$

$$SSA = n\left[\overline{(\overline{x}_{a_i})^2} - (\overline{x})^2\right]$$

$$= 27\left(\frac{11,87^2 + 16^2 + 21,11^2}{3} - 16,32^2\right) = 386,00$$

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

$$SSB = n\left[\overline{(\bar{x}_{b_j})^2} - (\bar{x})^2\right]$$

$$= 27\left(\frac{13{,}42^2 + 19{,}61^2 + 15{,}94^2}{3} - 16{,}32^2\right) = 174{,}32$$

SSAB = SS(AB) - SSA - SSB = 607,84 - 386 - 174,32 = 47,52

SSE = TSS - SS(AB) = 621,17 - 607,84 = 13,33

$$S_E^2 = \frac{SSE}{n-ab} = \frac{13{,}33}{27-3\cdot 3} = 0{,}74$$

Giả thuyết:

H_o: Không có tương tác giữa chính sách giá và cách thức bày hàng

H_1: Có tương tác giữa chính sách giá và cách thức bày hàng

Tiêu chuẩn kiểm định: $F_{AB} = \frac{S_{AB}^2}{S_E^2}$

Trong đó: $S_{AB}^2 = \frac{SSAB}{(a-1)(b-1)} = \frac{47{,}52}{(3-1)(3-1)} = 11{,}88$

$$\Rightarrow \quad F_{AB} = \frac{S_{AB}^2}{S_E^2} = \frac{11{,}88}{0{,}74} = 16{,}05$$

Tra bảng: $F_{(a-1)(b-1),n-ab,\alpha} = F_{4;18;0,05} = 2{,}93$

$F_{AB} > F_{(a-1)(b-1),n-ab,\alpha}$ => Bác bỏ H_o hay giữa chính sách giá và cách thức bày hàng tồn tại sự tương tác.

Giả thuyết: $H_o : \mu_{a1} = \mu_{a2} = \mu_{a3}$

$H_1 : \exists j \neq i$ mà $\mu_{ai} \neq \mu_{aj}$ (với j, i =1, 2, 3)

Tiêu chuẩn kiểm định: $F_A = \frac{S_A^2}{S_E^2}$

Trong đó: $S_A^2 = \frac{SSA}{a-1} = \frac{386{,}00}{3-1} = 193{,}00$

$$\Rightarrow \quad F_A = \frac{S_A^2}{S_E^2} = \frac{193}{0{,}74} = 260{,}68$$

Tra bảng: $F_{a-1;n-ab,\alpha} = F_{2;18;0,05} = 3{,}55$

$F_A > F_{a-1,n-ab,\alpha}$: Bác bỏ H_o, hay các chính sách giá có ảnh hưởng khác nhau đến doanh số.

Giả thuyết: $H_o : \mu_{b1} = \mu_{b2} = \mu_{b3}$

$H_1 : \exists i \neq j$ mà $\mu_{bj} \neq \mu_{bi}$ (với i, j =1, 2, 3)

Tiêu chuẩn kiểm định: $F_{AB} = \frac{S_{AB}^2}{S_E^2}$

Trong đó: $S_B^2 = \frac{SSB}{b-1} = \frac{174{,}32}{3-1} = 87{,}16$

$$\Rightarrow \quad F_B = \frac{S_B^2}{S_E^2} = \frac{87{,}16}{0{,}74} = 117{,}78$$

Tra bảng: $F_{b-1;n-ab,\alpha} = F_{2;18;0,05} = 3{,}55$

$F_B > F_{b-1,n-ab,\alpha}$: Bác bỏ H_o, hay các cách thức bày hàng có ảnh hưởng khác nhau đến doanh số.

9.3. Kiểm định bội (kiểm định từng cặp một)

Trong các phương pháp phân tích phương sai được trình bày ở trên, khi huỷ bỏ giả thuyết H_o ta biết được rằng tồn tại ít nhất một cặp Số trung bình của các tổng thể khác nhau. Tuy nhiên, ta không biết cụ thể những cặp tổng thể nào khác nhau. Kiểm định bội hay kiểm định từng cặp một sẽ giúp ta trả lời câu hỏi này.

Có nhiều phương pháp kiểm định bội khác nhau, hai phương pháp thường được dùng là phương pháp Tukey và phương pháp Student-Newman-Keuls.

9.3.1. Phương pháp Tukey

Nếu giả thuyết H_o về sự bằng nhau của tất cả các cặp Số trung bình tổng thể là đúng, thì:

$$\frac{\overline{X}_{max} - \overline{X}_{min}}{\sqrt{\frac{S_E^2}{m}}} \sim q_{r_1,r_2}$$

Trong đó:

* $\overline{X}_{max}$, $\overline{X}_{min}$: Số trung bình mẫu lớn nhất, nhỏ nhất,
* S_E^2 : Trung bình bình phương sai số trong phân tích phương sai,
* m : Kích thước mẫu đang so sánh,
* q_{r_1,r_2} : Phân phối Studentized Range với r_1, r_2 bậc tự do,
* r_1 : Số tổng thể đang muốn so sánh,

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

* r_2 : Bậc tự do của S_E^2 trong phân tích phương sai.

Do đó, *tiêu chuẩn kiểm định*: $$q = q_{r_1, r_2, \alpha}\sqrt{\frac{S_E^2}{m}}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):

* $\left|\bar{x}_i - \bar{x}_j\right| \geq q \Rightarrow \mu_i \neq \mu_j$

* $\left|\bar{x}_i - \bar{x}_j\right| < q \Rightarrow \mu_i = \mu_j$

Chú ý:

* Kiểm định bội chỉ cần tiến hành khi phân tích phương sai có đủ cơ sở bác bỏ giả thuyết H_o về sự bằng nhau của tất cả các tổng thể đang muốn kiểm định bội.

* Với phân tích phương sai một yếu tố S_E^2 chính là S_W^2.

* Nếu các n_i trong cặp mẫu so sánh là khác nhau thì: $m = \frac{2n_i n_j}{n_i + n_j}$.

Ví dụ: Kiểm định bội chi tiêu trung bình của du khách ở bốn thị trường trọng điểm đã xét đến ở ví dụ của mục phân tích phương sai một yếu tố với α = 0,05.

Kết quả phân tích phương sai trước đây cho biết:

$m = 8$, $s_W^2 = 54{,}26$, $r_1 = 4$, $r_2 = 28$

$\bar{x}_1 = 3{,}75$, $\bar{x}_2 = 10{,}63$, $\bar{x}_3 = 22{,}50$, $\bar{x}_4 = 12{,}00$

Tiêu chuẩn kiểm định:

$$q = q_{r_1, r_2, \alpha}\sqrt{\frac{S^2{}_W}{m}} = q_{4;28;0,05}\sqrt{\frac{54{,}26}{8}} = 3{,}87\sqrt{\frac{54{,}26}{8}} = 10{,}06$$

Do đó:

$\left|\bar{x}_1 - \bar{x}_2\right| = 6{,}88 < q \Rightarrow \mu_1 = \mu_2$

$\left|\bar{x}_1 - \bar{x}_3\right| = 18{,}75 > q \Rightarrow \mu_1 \neq \mu_3$

$\left|\bar{x}_1 - \bar{x}_4\right| = 8{,}25 < q \Rightarrow \mu_1 = \mu_4$

$\left|\bar{x}_2 - \bar{x}_3\right| = 11{,}87 > q \Rightarrow \mu_2 \neq \mu_3$

$\left|\bar{x}_2 - \bar{x}_4\right| = 1{,}37 < q \Rightarrow \mu_2 = \mu_4$

$\left|\bar{x}_3 - \bar{x}_4\right| = 10{,}5 > q \Rightarrow \mu_3 \neq \mu_4$

9.3.2. Phương pháp Student-Newman-Keuls

Theo phương pháp này, *cần sắp xếp các Số trung bình mẫu thành một dãy có thứ tự, lớn dần*. Khoảng cách giữa hai số trung bình muốn so sánh với nhau gọi là số bước (step) giữa chúng. *Số bước giữa hai số trung bình muốn so sánh cách nhau k Số trung bình ở giữa là: s = 2 + k.*

Nếu giả thuyết H_o về sự bằng nhau của tất cả các cặp Số trung bình tổng thể là đúng, thì:

$$\frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{S_E^2}{m}}} \sim q_{s, r_2}$$

Trong đó:

* $\bar{X}_i$, $\bar{X}_j$: Số trung bình mẫu của tổng thể i, j trong dãy số trung bình đã được sắp xếp;

* S_E^2 : Trung bình bình phương sai số trong phân tích phương sai;

* m : Kích thước mẫu đang so sánh;

* q_{s, r_2} : Phân phối Studentized Range với s, r_2 bậc tự do;

* s : Số bước giữa hai số trung bình μ_i, μ_j cần so sánh;

* r_2 : Bậc tự do của S_E^2 trong phân tích phương sai.

Do đó, chọn *tiêu chuẩn kiểm định :* $$q_s = q_{s, r_2, \alpha}\sqrt{\frac{S_E^2}{m}}$$

Ta kết luận với mức ý nghĩa α (xác suất sai lầm loại I):

* $\left|\bar{x}_i - \bar{x}_j\right| \geq q_s \Rightarrow \mu_i \neq \mu_j$

* $\left|\bar{x}_i - \bar{x}_j\right| < q_s \Rightarrow \mu_i = \mu_j$

Ví dụ: Kiểm định bội bốn cách thức nghỉ ngơi giữa ca đã xét đến ở ví dụ của mục phân tích phương sai hai yếu tố không lặp, với α = 0,05.

Kết quả phân tích phương sai trước đây cho biết:

$m = 5$, $S_E^2 = 0{,}89$, $r_2 = 12$

$\bar{x}_1 = 14{,}2$, $\bar{x}_2 = 16{,}4$, $\bar{x}_3 = 16{,}2$, $\bar{x}_4 = 13{,}6$

Sắp xếp các số trung bình mẫu theo độ lớn:

$\bar{x}_4$ $\bar{x}_1$ $\bar{x}_3$ $\bar{x}_2$

Tiêu chuẩn kiểm định:

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

$$q_s = q_{s,r_2,\alpha}\sqrt{\frac{S_E^2}{m}} = q_{s;12;0,05}\sqrt{\frac{0,89}{5}} = q_{s;12;0,05}\cdot 0,422$$

Lập bảng tiêu chuẩn kiểm định q_s theo số bước:

s	2	3	4
$q_{s;\ 12;\ 0,05}$	3,08	3,77	4,20
q_s	1,30	1,59	1,77

Do đó:

$$|\bar{x}_2 - \bar{x}_4| = 2,8 > q_4 \Rightarrow \mu_2 \neq \mu_4$$
$$|\bar{x}_3 - \bar{x}_4| = 2,6 > q_3 \Rightarrow \mu_3 \neq \mu_4$$
$$|\bar{x}_1 - \bar{x}_4| = 0,6 < q_2 \Rightarrow \mu_1 = \mu_4$$
$$|\bar{x}_2 - \bar{x}_1| = 2,2 > q_3 \Rightarrow \mu_2 \neq \mu_1$$
$$|\bar{x}_3 - \bar{x}_1| = 2,0 > q_2 \Rightarrow \mu_3 \neq \mu_1$$
$$|\bar{x}_2 - \bar{x}_3| = 0,2 < q_2 \Rightarrow \mu_2 = \mu_3$$

Kết quả kiểm định có thể được trình bày bằng hình vẽ nối các tổng thể bằng nhau như sau: $\underline{\mu_4 \quad \mu_1} \quad \underline{\mu_3 \quad \mu_2}$

BÀI TẬP

Chương 2

Bài 1. Chính quyền thành phố thực hiện khảo sát ý kiến người dân về dịch vụ công tại các cơ quan công quyền. Một câu hỏi trong phiếu khảo sát là: Xin vui lòng viết rõ những góp ý về dịch vụ công hiện nay cần phải hoàn thiện:..... . Trong 75 phiếu khảo sát thu về có 21 diễn đạt góp ý khác nhau cho câu hỏi trên và được mã hóa bằng các chữ cái từ A đến V với số phiếu trả lời tương ứng là số trong ngoặc như trong bảng sau:

A (9)	B (2)	C (1)	D (3)	E (1)	F (5)	G (1)
H (1)	I (4)	K (2)	L (6)	M (1)	N (8)	O (13)
P (3)	Q (1)	R (4)	S (2)	T (1)	U (1)	V (6)

Sau khi xem xét kỹ lưỡng các phiếu trả lời, chính quyền thành phố nhận thấy về cơ bản một số diễn đạt góp ý trong các mã hóa trên là gần giống nhau về nội dung như sau: Các diễn đạt góp ý K, O, Q, V là về thái độ công chức còn kém; các diễn đạt góp ý C, E, H, I, N, R, U là về thời gian xử lý kéo dài; các diễn đạt góp ý A, F, L, M, P là về qui trình thủ tục phức tạp; các diễn đạt góp ý B, D, S là về thời gian xếp hàng; các góp ý G, T là về trang thiết bị làm việc.

1. Hãy mô tả dữ liệu các góp ý về thái độ công chức còn kém của cuộc khảo sát trên bằng bảng phân phối tần suất phần trăm.
2. Hãy mô tả dữ liệu về 75 góp ý của cuộc khảo sát bằng bảng phân phối tần số.
3. Hãy mô tả dữ liệu về 75 góp ý của cuộc khảo sát trên bằng biểu đồ hình thanh.
4. Hãy mô tả dữ liệu về 75 góp ý của cuộc khảo sát trên trên bằng biểu đồ hình bánh (hình tròn).

Bài 2. Một công ty thực hiện khảo sát ý kiến người tiêu dùng về chất lượng quạt máy của họ. Một câu hỏi trong phiếu khảo sát là: Xin vui lòng viết rõ những góp ý về sản phẩm để chúng tôi hoàn thiện:..... . Trong 100 phiếu

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

khảo sát thu về có 24 diễn đạt góp ý khác nhau cho câu hỏi trên và được mã hóa bằng các chữ cái từ A đến Z với số phiếu tương ứng là số trong ngoặc như trong bảng sau:

A (1)	B (2)	C (1)	D (4)	E (2)	F (5)	G (3)	H (6)
I (4)	K (3)	L (2)	M (1)	N (5)	O (8)	P (2)	Q (16)
R (12)	S (1)	T (9)	U (3)	V (3)	X (2)	Y (1)	Z (4)

Sau khi xem xét kỹ lưỡng các phiếu trả lời, công ty thấy về cơ bản một số diễn đạt góp ý trong các mã hóa trên là gần giống nhau về nội dung như sau: Các góp ý A, D, Q là về hao phí điện năng còn cao; Các góp ý C, P, F, V, Z, I là về kiểu dáng chưa đẹp; các góp ý B, M, H, Y, K, T, O, X là về tiếng ồn còn lớn; các góp ý R, L là về màu sắc chưa nhiều; các góp ý E, N, G, U, S là về dịch vụ bảo hành còn kém.

1. Hãy mô tả dữ liệu các góp ý về hao phí điện năng của cuộc khảo sát trên bằng bảng phân phối tần suất phần trăm.
2. Hãy mô tả dữ liệu về 100 góp ý của cuộc khảo sát bằng bảng phân phối tần số.
3. Hãy mô tả dữ liệu về 100 góp ý của cuộc khảo sát trên bằng biểu đồ hình thanh.
4. Hãy mô tả dữ liệu về 100 góp ý của cuộc khảo sát trên trên bằng biểu đồ hình bánh (hình tròn).

Bài 3. Có dữ liệu về thu nhập của 50 người tiêu dùng ở thị trường TP Huế được chọn ngẫu nhiên như sau:

Đơn vị tính: Triệu đồng

5,48	3,91	1,42	1,60	2,30	4,10	9,00	6,50	6,50	8,10
5,55	3,92	1,42	1,66	2,45	4,50	9,50	7,20	7,40	8,40
5,57	3,95	1,42	1,84	2,68	5,60	9,00	7,50	7,50	8,80
5,65	3,98	1,42	1,95	3,50	5,80	9,00	7,70	8,00	9,10
5,85	4,25	1,42	1,98	3,80	6,10	10,00	8,10	8,60	9,50

1. Hãy mô tả dữ liệu trên bằng phân tổ đều, lập bảng tần số và cho nhận xét.
2. Hãy mô tả dữ liệu đã được phân tổ bằng biểu đồ phân phối.
3. Hãy mô tả dữ liệu trên bằng biểu đồ cành và lá.
4. Hãy mô tả bằng Mốt dữ liệu trên và cho nhân xét.
5. Hãy mô tả dữ liệu trên bằng số trung bình.
6. Hãy mô tả dữ liệu trên bằng phương sai.
7. Hãy mô tả dữ liệu trên bằng các Tứ phân vị và biểu diễn trên biểu đồ hộp.

Bài 4. Có dữ liệu về tuổi nghề của 40 công nhân:

3	1	5	7	15	27	33	21
4	2	5	7	16	28	34	24
4	2	5	8	19	30	36	25
4	3	6	10	19	30	38	25
4	3	7	12	21	32	40	25

1. Hãy mô tả dữ liệu trên bằng phân tổ đều, lập bảng tần số và cho nhận xét.
2. Hãy mô tả dữ liệu đã được phân tổ bằng biểu đồ phân phối.
3. Hãy lập bảng phân phối dữ liệu đã được phân tổ theo tần suất, tần suất %, tần số tích lũy, tần suất tích lũy và cho nhận xét;
4. Hãy vẽ biểu đồ hình cung mô tả phân phối dữ liệu đã được phân tổ theo tần suất tích lũy và cho nhận xét.
5. Hãy mô tả dữ liệu trên bằng biểu đồ cành và lá và cho nhận xét.
6. Hãy mô tả dữ liệu trên bằng các chỉ tiêu (số đo) vị trí trung tâm (khuynh hướng hội tụ), so sánh các kết quả và cho nhận xét..
7. Hãy mô tả dữ liệu trên bằng phương sai.
8. Hãy mô tả dữ liệu trên bằng các Tứ phân vị và biểu diễn trên biểu đồ hộp.

Bài 5. Có dữ liệu về thu nhập (Triệu đồng) của 70 khách hàng tại một siêu thị được chọn ngẫu nhiên và sắp xếp như sau:

1,52	2,01	3,12	4,60	7,30	10,10	81,00	25,50	31,10	41,50
1,60	2,12	3,42	4,66	7,45	10,50	83,60	26,20	32,40	43,40
1,72	2,25	3,42	4,84	7,68	15,60	84,00	27,50	34,80	61,50
1,82	2,58	3,72	4,95	7,70	18,50	85,00	27,70	35,10	65,00
1,85	2,75	4,22	6,08	7,80	21,10	92,00	28,10	36,50	61,60
1,90	3,03	4,30	6,25	7,92	22,40	95,00	30,00	38,00	71,00
1,94	3,08	4,45	6,34	7,98	24,60	98,00	30,50	40,00	75,50

1. Hãy mô tả dữ liệu trên bằng phân tổ đều, lập bảng tần số và cho nhận xét.
2. Hãy mô tả dữ liệu đã được phân tổ bằng biểu đồ phân phối.
3. Hãy lập bảng phân phối dữ liệu đã được phân tổ theo tần suất, tần suất %, tần số tích lũy, tần suất tích lũy và cho nhận xét;
4. Hãy vẽ biểu đồ hình cung mô tả phân phối dữ liệu đã được phân tổ theo tần suất tích lũy và cho nhận xét.
5. Hãy mô tả dữ liệu trên bằng biểu đồ cành và lá và cho nhận xét.
6. Hãy mô tả dữ liệu đã được phân tổ ở câu 1 bằng số trung bình.

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Bài 6. Có tài liệu về tuổi nghề và năng suất lao động của 15 công nhân được chọn ngẫu nhiên ở một doanh nghiệp như sau:

Công nhân	Tuổi nghề	Loại sức khỏe	Năng suất (SP/giờ)
1	1	C	9
2	2	B	11
3	2	C	10
4	3	A	12
5	5	C	12
6	5	C	11
7	6	B	13
8	7	B	12
9	8	B	14
10	10	A	14
11	12	A	16
12	14	A	16
13	16	A	17
14	18	B	15
15	23	A	18

1. Hãy phân tổ kết hợp các công nhân theo hai tiêu thức loại sức khỏe và năng suất (hai tổ đều), lập bảng phân phối chéo và cho nhận xét.
2. Dựa trên bảng chéo câu 1, hãy mô tả kết cấu công nhân theo loại sức khỏe bằng biểu đồ thích hợp.
3. Dựa trên bảng chéo câu 1, lập bảng chéo tỉ lệ phần trăm theo cột và cho nhận xét.
4. Dựa trên bảng chéo câu 1, lập bảng chéo tỉ lệ phần trăm theo dòng và cho nhận xét.
5. Dựa trên bảng chéo ở câu 1 vẽ biểu đồ nhiều thanh cạnh nhau và cho nhận xét.
6. Dựa trên bảng chéo tỉ lệ phần trăm theo cột ở câu 3 vẽ biểu đồ nhiều thanh chồng nhau và cho nhận xét.
7. Hãy phân tổ kết hợp các công nhân theo hai tiêu thức tuổi nghề (hai tổ đều) và năng suất (hai tổ đều) và lập bảng phân phối chéo.
8. Dựa trên bảng chéo câu 6, lập bảng chéo tỉ lệ phần trăm theo cột và cho nhận xét.
9. Dựa trên bảng chéo câu 6, lập bảng chéo tỉ lệ phần trăm theo dòng và cho nhận xét.
10. Dựa trên bảng chéo ở câu 6 vẽ biểu đồ nhiều thanh cạnh nhau và cho nhận xét.
11. Dựa trên bảng chéo tỉ lệ phần trăm theo dòng ở câu 8 vẽ biểu đồ nhiều thanh chồng nhau và cho nhận xét.
12. Vẽ đồ thị phân tán và đường xu hướng mô tả liên hệ giữa hai tiêu thức tuổi nghề và năng suất.
13. Hãy xác định số trung vị của năng suất.
14. Hãy mô tả năng suất bằng các chỉ tiêu vị trí trung tâm (khuynh hướng hội tụ), so sánh các kết quả và cho nhận xét.
15. Hãy mô tả năng suất bằng các chỉ tiêu độ phân tán.

.

Bài 7. Có dữ liệu về 100 khách hàng được chọn ngẫu nhiên tại một nhà hàng như sau:

Số thành viên trong gia đình	Số người
1	10
2	20
3	50
4	15
5	5

1. Hãy mô tả dữ liệu trên bằng biểu đồ phân phối;
2. Hãy mô tả dữ liệu trên bằng số trung bình;
3. Hãy mô tả dữ liệu trên bằng số trung vị;
4. Hãy mô tả dữ liệu trên bằng Mốt;
5. Hãy mô tả dữ liệu trên bằng các Tứ phân vị và biểu diễn trên biểu đồ hộp.

Bài 8. Có dữ liệu phân tổ 360 người tiêu dùng được chọn ngẫu nhiên theo thu nhập hàng tháng như trang sau.

Thu nhập (triệu đồng/tháng)	Số người
Dưới 0,5	10
0,5 - 1,0	50
1,0 - 2,0	235
2,0 - 4,0	45
4,0 - 8,0	14
8,0 trở lên	6

1. Hãy mô tả dữ liệu trên bằng biểu đồ thích hợp;
2. Hãy tính thu nhập trung bình một tháng của dữ liệu mẫu trên.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

3. Hãy tính phương sai dữ liệu mẫu trên.

Bài 9. Có tình hình sản xuất một loại sản phẩm ở một doanh nghiệp như sau:

Phân xưởng	Giá thành (1000đ/SP)		Sản lượng (SP)	
	Tháng 1	Tháng 2	Tháng 1	Tháng 2
1	130	128	150	220
2	128	126	220	240
3	132	130	380	350

Yêu cầu: Tính giá thành trung bình có trọng số một sản phẩm ở mỗi tháng.

Bài 10. Có tình hình sản xuất một loại sản phẩm ở một doanh nghiệp như sau:

Phân xưởng	Giá thành (1000đ /SP)		Chi phí sản xuất (triệu đồng)	
	Tháng 1	Tháng 2	Tháng 1	Tháng 2
1	30	28	100	220
2	28	42	200	140
3	42	30	300	350

Yêu cầu: Hãy tính giá thành trung bình có trọng số một sản phẩm ở mỗi tháng.

Bài 11. Có dữ liệu về một danh mục đầu tư như sau:

Loại chứng khoán	Số lượng	Đơn giá (nghìn đồng)	Lợi suất (%)
A	700	40	5,04
B	500	60	6,60
C	1200	50	9,60
D	200	70	3,50
E	3000	30	18,00

Yêu cầu: Hãy xác định lợi suất trung bình của danh mục.

Bài 12. Có dữ liệu về một danh mục đầu tư như sau:

Loại chứng khoán	Tỉ trọng vốn đầu tư (%)	Lợi suất (%)
A	40	5,04
B	10	6,60
C	30	9,60
D	15	3,50
E	5	18,00

Yêu cầu: Hãy xác định lợi suất trung bình của danh mục.

Bài 13. Một doanh nghiệp có hai cửa hàng cùng bán ra chỉ một loại hàng. Năm 2015, cửa hàng thứ nhất có doanh số bán là 50 triệu đồng và cửa hàng thứ hai có doanh số bán là 80 triệu đồng. Năm 2016, cửa hàng thứ nhất chiếm 30% tổng lượng bán của doanh nghiệp. Đơn giá bán của cửa hàng thứ nhất năm 2015 là 2,5 (1000đ/SP), năm 2016 là 2,6 (1000đ/SP). Đơn giá bán của cửa hàng thứ hai năm 2015 là 2,4 (1000đ/SP), năm 2016 là 2,5 (1000đ/SP).

1. Xác định giá bán trung bình có trọng số của doanh nghiệp ở năm 2015.
2. Xác định giá bán trung bình có trọng số của doanh nghiệp ở năm 2016.

Bài 14. Để nghiên cứu chi tiêu của các hộ gia đình, người ta chia các hộ gia đình của một thành phố làm ba vùng: Ngoại ô gồm 500 hộ, ven đô gồm 1000 hộ, trung tâm gồm 2000 hộ. Sau đó người ta chọn ngẫu nhiên 7 hộ ngoại ô, 20 hộ ven đô và 30 hộ trung tâm. Dữ liệu về chi tiêu hàng tháng (triệu đồng) của các hộ thu được như sau:

Các hộ ngoại ô:

1,8	2,4	2,8	3,0	4,4	5,3	5,0

Các hộ ven đô:

3,5	3,5	3,5	2,5	3,5	4,2	4,2	4,2	4,2	4,2
5,5	5,5	5,5	5,5	5,5	8,3	8,3	5,5	8,5	5,5

Các hộ trung tâm:

3,6	4,0	4,4	4,4	4,7	5,0	5,3	5,5	5,6	5,9
6,0	6,1	6,3	6,3	6,7	6,8	6,9	7,2	7,2	7,4
7,5	7,5	7,5	7,7	8,0	8,4	8,8	9,2	9,7	9,5

1. Hãy mô tả dữ liệu trên bằng biểu đồ hộp cho từng khu vực, so sánh các kết quả và cho nhận xét;
2. Hãy mô tả dữ liệu trên bằng các chỉ tiêu đo lường vị trí trung tâm (khuynh hướng hội tụ) cho từng khu vực, so sánh các kết quả và cho nhận xét;
3. Hãy xác định các chỉ tiêu đo lường độ phân tán cho khu vực ngoại ô.
4. Lập bảng phân phối tần số cho các hộ ven đô. Dựa vào bảng này tính độ lệch tuyệt đối trung bình và phương sai.
5. Hãy so sánh độ phân tán chi tiêu giữa khu vực ngoại ô và khu vực ven đô, cho nhận xét.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Bài 15. Có dữ liệu về tuổi nghề và năng suất lao động của 8 công nhân được chọn ngẫu nhiên ở một doanh nghiệp như sau:

Công nhân	Tuổi nghề	Năng suất (SP/giờ)
1	3	12
2	2	11
3	18	14
4	5	13
5	12	16
6	2	10
7	6	15
8	9	14

1. Xác định Hệ số Kurtosis, Hệ số Skewnes của tuổi nghề. Cho nhận xét về hình dáng phân phối của tiêu thức này.
2. Mô tả mối liên hệ tương quan giữa tuổi nghề và năng suất bằng Hiệp phương sai.
3. Mô tả mối liên hệ tương quan giữa tuổi nghề và năng suất bằng hệ số tương quan Pearson.
4. Mô tả mối liên hệ tương quan giữa tuổi nghề và năng suất bằng hệ số tương quan hạng Spearman.

Bài 16. Có dữ liệu về chi phí quảng cáo và tỉ suất lợi nhuận trên vốn của 10 doanh nghiệp được chọn ngẫu nhiên ở một thành phố như trang sau.

Doanh nghiệp	Chi phí quảng cáo (triệu đồng)	Tỉ suất lợi nhuận trên vốn (%)
1	230	12
2	450	11
3	370	15
4	800	13
5	540	11
6	120	16
7	750	14
8	920	12
9	400	10
10	380	15

1. Xác định Hệ số Kurtosis, Hệ số Skewnes của tỉ suất lợi nhuận trên vốn. Cho nhận xét về hình dáng phân phối của tiêu thức này.
2. Mô tả mối liên hệ tương quan giữa chi phí quảng cáo và tỉ suất lợi nhuận trên vốn bằng Hiệp phương sai.
3. Mô tả mối liên hệ tương quan giữa chi phí quảng cáo và tỉ suất lợi nhuận trên vốn bằng hệ số tương quan Pearson.
4. Mô tả mối liên hệ tương quan giữa chi phí quảng cáo và tỉ suất lợi nhuận trên vốn bằng hệ số tương quan hạng Spearman.

Bài 17. Để nghiên cứu mối liên hệ giữa giới tính và loại bao bì của một loại sản phẩm, một doanh nghiệp chọn ngẫu nhiên 100 khách hàng gồm 60 nam và 40 nữ để điều tra xem họ thích chọn loại bao bì nào. Kết quả điều tra được phân tổ như trong bảng sau:

Giới tính	Loại bao bì lựa chọn			Cộng
	A	B	C	
Nam	33	20	7	60
Nữ	8	10	22	40
Cộng	41	30	29	100

Hãy mô tả mối liên hệ giữa giới tính và loại bao bì bằng Hệ số Cramer, Hệ số liên hợp.

Bài 18. Có dữ liệu về 140 doanh nghiệp được chọn ngẫu nhiên ở một thành phố và được phân tổ kết hợp theo loại hình kinh doanh và tỉ suất lợi nhuận trên vốn như sau:

Loại hình kinh doanh	Tỉ suất lợi nhuận trên vốn (%)			Cộng
	5-10	10-15	15-20	
Sản suất	20	60	6	86
Dịch vụ	5	30	19	54
Cộng	25	90	25	140

Hãy mô tả mối liên hệ giữa loại hình kinh doanh và tỉ suất lợi nhuận trên vốn bằng Hệ số Cramer, Hệ số liên hợp.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Bài 19. Có dữ liệu về 140 doanh nghiệp được chọn ngẫu nhiên ở một thành phố và được phân tổ kết hợp theo quy mô và tỉ suất lợi nhuận trên vốn như sau:

Hãy mô tả mối liên hệ giữa quy mô và tỉ suất lợi nhuận trên vốn bằng Hệ số Gamma, Hệ số Kendall-Tau A.

Quy mô	Tỉ suất lợi nhuận trên vốn (%)			Cộng
	5-10	10-15	15-20	
Nhỏ	20	60	6	86
Vừa	5	30	19	54
Cộng	25	90	25	140

Bài 20. Có dữ liệu về 110 doanh nghiệp được chọn ngẫu nhiên ở một thành phố và được phân tổ kết hợp theo tỉ suất lợi nhuận và xếp loại tín dụng như sau:

Tỉ suất lợi nhuận (%)	Xếp loại tín dụng			Cộng
	Kém	Trung bình	Khá	
Dưới 10	18	10	7	35
10-14	8	15	12	35
Trên 14	5	9	26	40
Cộng	31	34	45	110

Hãy mô tả mối liên hệ giữa xếp loại tín dụng và tỉ suất lợi nhuận trên vốn bằng Hệ số Gamma, Hệ số Kendall-Tau A.

Chương 3

Bài 21. Giả sử, trong 5000 chứng từ của doanh nghiệp có 1000 chứng từ không hợp lệ.

Yêu cầu:

1. Chọn ngẫu nhiên một chứng từ. Hãy viết phân phối xác suất về tính hợp lệ của chứng từ đó. *(gợi ý: xem ý nghĩa vận dụng của phân phối Bernoulli)*
2. Chọn lặp, ngẫu nhiên 10 chứng từ. Tìm xác suất có 2 chứng từ không hợp lệ trở lại. *(gợi ý: xem ý nghĩa vận dụng của phân phối nhị thức)*
3. Chọn lặp, ngẫu nhiên 12 chứng từ. Tìm xác suất có 10 chứng từ không hợp lệ trở lên. *(gợi ý: xem ý nghĩa vận dụng của phân phối nhị thức)*

Bài 22. Một công ty muốn kiểm tra hiểu biết của khách hàng về một tính năng sản phẩm mà công ty đưa vào sản phẩm đã được một thời gian. 7 câu hỏi về tính năng (mỗi câu có sẵn 5 lựa chọn trả lời trong đó chỉ có một lựa chọn trả lời đúng) đã được đặt ra cho 100 khách hàng được chọn ngẫu nhiên.

1. Lập phân phối xác suất cho số câu trả lời đúng của một khách hàng nếu khách hàng này hoàn toàn không biết gì về tính năng công ty đang khảo sát mà chỉ trả lời ngẫu nhiên. *(gợi ý: xem vận dụng thực tiễn của phân phối nhị thức, với 7 phép thử ngẫu nhiên cho một khách hàng)*
2. Lập phân phối tần số số khách hàng trả lời đúng theo số câu trả lời đúng nếu 100 khách hàng này hoàn toàn không biết gì về tính năng công ty đang khảo sát mà chỉ trả lời ngẫu nhiên. *(gợi ý: dựa trên kết quả câu 1 và số khách tham gia)*

Bài 23. Một lô sản phẩm nhập khẩu gồm 5000 sản phẩm. Trọng lượng trung bình một sản phẩm là 5,5kg. Độ lệch chuẩn trọng lượng sản phẩm là 0,8kg.

1. Chọn lặp, ngẫu nhiên 50 sản phẩm. Hãy viết phân phối xác suất của trọng lượng trung bình một sản phẩm được chọn. *(gợi ý: xem ý nghĩa vận dụng của định lý giới hạn trung tâm)*
2. Tìm xác suất để trọng lượng trung bình của 40 sản phẩm được chọn lặp ngẫu nhiên có trọng lượng từ 5,8 kg trở lên. *(gợi ý: tương tự câu 1 rồi chuyển đổi từ phân phối chuẩn sang phân phối chuẩn tắc để tìm xác suất đuôi phải)*
3. Tìm xác suất để trọng lượng trung bình của 60 sản phẩm được chọn lặp ngẫu nhiên có trọng lượng từ 5,4 kg trở xuống. *(gợi ý: tương tự câu 2 nhưng tìm xác suất đuôi trái = 1 - xác suất đuôi phải).*
4. Tìm xác suất để trọng lượng trung bình của 80 sản phẩm được chọn lặp ngẫu nhiên có trọng lượng trong khoảng từ 5,3 đến 5,8 kg. *(gợi ý: tương tự câu 2 tìm xác suất đuôi phải của 5,3 rồi trừ cho xác suất đuôi phải của 5,8)*

Chương 4

Bài 24. Với dữ liệu ở Bài 3, hãy thực hiện các công việc sau:

1. Ước lượng điểm thu nhập trung bình của người tiêu dùng thị trường TP Huế.
2. Ước lượng điểm Tỉ lệ người tiêu dùng thị trường TP Huế có thu nhập trong khoảng từ 5 đến 8 triệu.
3. Ước lượng điểm phương sai thu nhập của người tiêu dùng thị trường TP Huế.
4. Ước lượng thu nhập trung bình của người tiêu dùng thị trường TP Huế với độ tin cậy 95%.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

5. Ước lượng thu nhập trung bình của người tiêu dùng thị trường TP Huế với độ tin cậy 95% biết rằng phương sai thu nhập của toàn bộ người tiêu dùng ở thị trường TP Huế là 8,4.
6. Ước lượng Tỉ lệ người tiêu dùng thị trường TP Huế có thu nhập trong khoảng từ 5 đến 8 triệu với độ tin cậy 99%.
7. Ước lượng phương sai thu nhập của người tiêu dùng thị trường TP Huế với độ tin cậy 95%.
8. Xác định kích thước mẫu cần điều tra nếu cần ước lượng Thu nhập trung bình với biên sai số không vượt quá 0,5 triệu/người và độ tin cậy 99%.
9. Xác định kích thước mẫu cần điều tra nếu cần ước lượng Tỉ lệ người tiêu dùng thị trường TP Huế có thu nhập trong khoảng từ 5 đến 8 triệu với biên sai số không vượt quá 0,04 và độ tin cậy 95%.

Bài 25. Với dữ liệu ở Bài 7, hãy thực hiện các công việc sau:
1. Ước lượng Số thành viên trong gia đình trung bình của khách hàng của nhà hàng với độ tin cậy 99%. Biết rằng phương sai số thành viên của toàn bộ khách hàng của nhà hàng là 0,81.
2. Ước lượng Tỉ lệ khách hàng của nhà hàng có số thành viên trong gia đình từ 3 người trở lên với độ tin cậy 95%.
3. Ước lượng phương sai Số thành viên trong gia đình của khách hàng của nhà hàng với độ tin cậy 99%.
4. Xác định kích thước mẫu cần điều tra nếu cần ước lượng Số thành viên trong gia đình trung bình với sai số biên là 0,2 người và độ tin cậy 95%.
5. Xác định kích thước mẫu nếu cần ước lượng Tỉ lệ khách hàng có số thành viên từ 3 người trở lên với sai số biên 1% và độ tin cậy 99%.

Bài 26. Với dữ liệu ở Bài 14, hãy thực hiện các công việc sau:
1. Ước lượng chi tiêu trung bình một hộ của khu vực ngoại ô thành phố với độ tin cậy 95%.
2. Ước lượng Tỉ lệ hộ có chi tiêu từ 5 triệu trở lên của khu vực trung tâm thành phố với độ tin cậy 99%.
3. Ước lượng phương sai chi tiêu của các hộ ở khu vực ven đô thành phố với độ tin cậy 95%.
4. Xác định kích thước mẫu cần điều tra cho khu vực ven đô thành phố nếu cần ước lượng chi tiêu trung bình mỗi hộ với biên sai số không vượt quá 0,4 triệu/người và độ tin cậy 95%.
5. Xác định kích thước mẫu cần điều tra cho khu vực trung tâm thành phố nếu cần ước lượng Tỉ lệ hộ có chi tiêu từ 5 triệu trở lên với biên sai số không vượt quá 0,02 và độ tin cậy 95%.

Chương 5

Bài 27. Với dữ liệu ở Bài 3, yêu cầu:
1. Hãy kiểm định giả thuyết cho rằng thu nhập trung bình của người tiêu dùng thị trường TP Huế là không dưới 6 triệu đồng với $\alpha = 0,01$;
2. Hãy kiểm định giả thuyết cho rằng thu nhập trung bình của người tiêu dùng thị trường TP Huế là không dưới 6 triệu đồng với $\alpha = 0,01$ biết rằng phương sai thu nhập của toàn bộ người tiêu dùng thị trường TP Huế là 8,4.
3. Hãy kiểm định giả thuyết cho rằng Tỉ lệ người tiêu dùng thị trường TP Huế có thu nhập từ 5 đến 7 triệu là không dưới 40% với $\alpha = 0,01$.
4. Hãy kiểm định giả thuyết cho rằng phương sai thu nhập của người tiêu dùng thị trường TP Huế là không dưới 7 với $\alpha = 0,05$. Biết rằng thu nhập của người tiêu dùng thị trường TP Huế có phân phối chuẩn.

Bài 28. Một doanh nghiệp cam kết với khách hàng chiều dài sản phẩm trung bình là 5 mm. Để kiểm tra cam kết này khách hàng chọn ngẫu nhiên 22 sản phẩm để đo. Kết quả như sau:

Chiều dài sản phẩm (mm)	Số sản phẩm
4,9	2
5,0	4
5,1	10
5,2	6

Biết rằng, chiều dài sản phẩm có phân phối chuẩn.
1. Hãy kiểm định cam kết trên với $\alpha = 0,01$.
2. Hãy kiểm định cam kết trên với $\alpha = 0,05$. Biết rằng phương sai chiều dài sản phẩm là 0,007

Bài 29. Tiêu chuẩn chất lượng đặt ra cho một loại sản phẩm của nhà máy là tuổi thọ trung bình một sản phẩm từ 5 nghìn giờ sử dụng trở lên. Để kiểm tra, người ta chọn ngẫu nhiên hoàn toàn 35 sản phẩm. Kết quả thu được như sau:

Tuổi thọ (nghìn giờ)	Số sản phẩm
4,5	2
4,8	8
4,9	18
5,0	5
5,1	2

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

1. Hãy kiểm định tiêu chuẩn chất lượng nói trên với mức ý nghĩa $\alpha = 0,01$.
2. Biết rằng phương sai tuổi thọ là 0,012. Hãy kiểm định tiêu chuẩn chất lượng nói trên với mức ý nghĩa $\alpha = 0,05$.

Bài 30. Theo hợp đồng ký kết với khách hàng, trọng lượng trung bình 1 sản phẩm nằm trong khoảng từ 9,9 đến 10,1 kg/SP. Để kiểm tra cam kết này, người ta chọn ngẫu nhiên 50 sản phẩm. Kết quả thu được như sau.

Trọng lượng (kg)	Số sản phẩm
9,4	3
9,5	10
9,8	25
9,9	8
10,0	4

Hãy kiểm định cam kết trên với $\alpha = 0,05$.

Bài 31. Một công nhân cho rằng tỉ lệ thành phẩm của chiếc máy hiện nay là từ 98% trở xuống. Để kiểm tra nhận định này, người công nhân sản xuất thử 5000 sản phẩm thì thấy có 10 sản phẩm hỏng. Hãy kiểm định nhận định trên với $\alpha = 0,01$.

Bài 32. Tiêu chuẩn chất lượng của một doanh nghiệp đặt ra là tỉ lệ sản phẩm có khuyết tật là 0,12%. Để kiểm định tiêu chuẩn này, người ta chọn ngẫu nhiên 2500 sản phẩm để kiểm tra. Kết quả cho thấy có 4 sản phẩm bị khuyết tật. Hãy kiểm định giả thiết trên với $\alpha = 0,01$.

Bài 33. Cam kết hợp đồng đặt ra là phương sai trọng lượng sản phẩm không quá 0,4. Để kiểm tra, người ta chọn ngẫu nhiên 78 sản phẩm và thu được kết quả như trong bảng sau.

Hãy kiểm định cam kết trên với $\alpha = 0,01$. Biết rằng trọng lượng sản phẩm có phân phối chuẩn.

Trọng lượng (kg)	Số sản phẩm
2,40	3
2,45	5
2,50	46
2,55	11
2,60	8
2,65	5

Bài 34. Một công nhân cho rằng độ đồng đều về trọng lượng của sản phẩm của doanh nghiệp hiện nay đã giảm xuống dưới mức cho phép. Cụ thể là phương sai trọng lượng sản phẩm hiện nay đã vượt quá 0,2 kg. Để kiểm tra, công nhân này chọn ngẫu nhiên 30 sản phẩm để đo lường. Kết quả như sau:

Trọng lượng (kg)	Số sản phẩm
10,0	3
10,2	5
10,3	6
10,4	8
10,5	6
10,6	2

Hãy kiểm định nhận định trên với $\alpha = 0,01$. Biết rằng trọng lượng sản phẩm có phân phối chuẩn.

Bài 35. Một giám đốc khách hàng cho rằng nếu chi phí thêm cho bao bì 1000 đ/sp thì có thể bán hàng với giá cao hơn giá cũ từ 3000đ/sp trở lên. Để kiểm tra nhận định này, vị giám đốc này tiến hành thử nghiệm trên 30 khách hàng được chọn ngẫu nhiên. Giá mà các khách hàng này chấp nhận mua với bao bì cũ (BBC) và với bao bì mới (BBM) như trong bảng sau (nghìn đồng/sp). Biết rằng phân phối giá mua của khách hàng có phân phối xấp xỉ chuẩn.

1. Hãy kiểm định nhận định trên với $\alpha = 0,01$.
2. Hãy ước lượng sự khác biệt giá trung bình BBC và BBM với độ tin cậy 95%.

Khách hàng	Giá mua BBC	Giá mua BBM	***Khách hàng***	Giá mua BBC	Giá mua BBM	***Khách hàng***	Giá mua BBC	Giá mua BBM
1	50	55	***11***	50	54	***21***	46	48
2	48	50	***12***	49	58	***22***	51	53
3	52	58	***13***	48	51	***23***	54	54
4	49	51	***14***	51	49	***24***	48	49
5	51	55	***15***	49	57	***25***	49	56
6	53	56	***16***	49	54	***26***	52	58
7	47	50	***17***	50	56	***27***	45	48
8	52	56	***18***	51	51	***28***	49	52
9	50	52	***19***	53	50	***29***	52	54
10	48	52	***20***	47	53	***30***	51	55

Bài 36. Người ta cho rằng phương pháp sản xuất X có chi phí tiền lương cao hơn phương pháp sản xuất Y từ 50đ/SP trở lên. Để kiểm tra người ta chọn

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

ngẫu nhiên 16 công nhân để thử nghiệm 2 phương pháp sản xuất này. Kết quả như trong bảng sau. Biết rằng phân phối chi phí lương của các công nhân tuân theo phân phối chuẩn.

Công nhân	Chi phí lương X (1000đ/SP)	Chi phí lương Y (1000đ/SP)	***Công nhân***	Chi phí lương X (1000đ/SP)	Chi phí lương Y (1000đ/SP)
1	5,1	5,0	***9***	5,8	5,5
2	6,0	5,8	***10***	6,1	5,9
3	5,8	5,5	***11***	6,3	6,1
4	5,4	5,3	***12***	5,2	5,4
5	5,8	5,9	***13***	5,6	5,2
6	5,0	5,1	***14***	5,9	5,7
7	5,2	5,0	***15***	6,2	6,0
8	5,5	5,3	***16***	6,0	5,7

1. Hãy kiểm định nhận định trên với $\alpha = 0,01$.
2. Hãy ước lượng sự khác biệt về chi phí sản xuất của hai phương pháp sản xuất với độ tin cậy 95%.

Bài 37. Một nhà cung cấp giới thiệu 2 kiểu thiết bị sản xuất cùng một loại sản phẩm cho khách hàng. Nhà cung cấp cho rằng, mặc dù kiểu máy 1 đắt hơn kiểu máy 2, tuy nhiên kiểu máy 1 cho phép tiết kiệm bình quân so với máy 2 trên 1 sản phẩm từ 0,1kg nguyên liệu trở lên. Để kiểm tra người mua sản xuất thử 25 sản phẩm trên máy 1 và 32 sản phẩm trên máy 2. Mức hao phí nguyên liệu cho 1 sản phẩm như sau:

Máy 1		Máy 2	
Mức hao phí NL (kg)	Số sản phẩm	Mức hao phí NL (kg)	Số sản phẩm
12,0	2	12,1	5
12,1	3	12,2	5
12,2	8	12,3	10
12,3	6	12,4	7
12,4	3	12,5	5
12,5	3		

1. Hãy kiểm định nhận định của nhà cung cấp trên với $\alpha = 0,01$. Biết rằng lượng nguyên vật liệu tiêu hao trên mỗi máy có phân phối chuẩn và phương sai hai máy lần lượt là 0,018 và 0,021. Sự khác biệt về mức hao phí nguyên vật liệu của hai máy là bao nhiêu với độ tin cậy 95%.
2. Hãy kiểm định nhận định của nhà cung cấp trên với $\alpha = 0,01$. Biết rằng lượng nguyên vật liệu tiêu hao trên mỗi máy có phân phối chuẩn. Sự khác biệt về mức hao phí nguyên vật liệu của hai máy là bao nhiêu với độ tin cậy 99%.
3. Kiểm định giả thuyết về sự bằng nhau của phương sai mức hao phí NL của hai máy trên với mức ý nghĩa $\alpha = 0,01$.

Bài 38. Người ta cho rằng chi phí điện năng cho một sản phẩm của máy X hiện đã lớn hơn máy Y từ 100đ/SP trở lên. Để kiểm tra nhận định này, người ta sản xuất thử 35 sản phẩm trên mỗi máy. Kết quả về chi phí điện năng như trong bảng trang sau.

Máy X		Máy Y	
Chi phí điện (1000đ/SP)	Số sản phẩm	Chi phí điện (1000đ/SP)	Số sản phẩm
4,8	1	4,6	2
4,9	5	4,7	5
5,0	8	4,8	7
5,1	12	4,9	13
5,2	6	5,0	6
5,3	3	5,1	2

1. Hãy kiểm định nhận định trên với mức ý nghĩa $\alpha = 0,01$. Biết rằng chi phí điện năng của 2 máy tuân theo phân phối chuẩn và phương sai hai máy lần lượt là 0,014 và 0,012. Sự khác biệt về chi phí điện năng của máy X và máy Y là bao nhiêu với độ tin cậy 95%.
2. Hãy kiểm định nhận định trên với mức ý nghĩa $\alpha = 0,01$. Biết rằng chi phí điện năng của 2 máy tuân theo phân phối chuẩn. Sự khác biệt về chi phí điện năng của máy X và máy Y là bao nhiêu với độ tin cậy 90%.
3. Hãy kiểm định giả thuyết về sự bằng nhau của phương sai chi phí điện năng của hai máy trên với mức ý nghĩa $\alpha = 0,01$.

Bài 39. Cũng thị trường TP Huế như ở bài 3, dữ liệu điều tra một mẫu ngẫu nhiên 40 người tiêu dùng 3 năm sau đó cho kết quả như trong bảng sau:

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Đơn vị tính: Triệu đồng

4,10	2,42	3,60	6,30	7,91	9,00	9,48	10,50	11,30	12,10
4,50	2,42	4,66	6,45	8,92	9,20	9,55	10,20	11,40	12,40
5,60	3,42	4,84	7,68	8,95	9,30	9,57	10,50	11,50	13,80
5,80	3,42	4,95	7,50	8,98	9,30	9,65	11,10	12,00	14,10

1. Hãy kiểm định nhận định cho rằng thu nhập trung bình của người tiêu dùng thị trường TP Huế đã tăng ít nhất 2,5 triệu đồng sau 3 năm với $\alpha = 0,01$.
2. Hãy ước lượng sự khác biệt về thu nhập trung bình của người tiêu dùng thị trường TP Huế hiện nay so với 3 năm trước với độ tin cậy 95%.
3. Hãy kiểm định nhận định cho rằng tỉ lệ người tiêu dùng thị trường TP Huế có thu nhập trên 8 triệu đã tăng rõ rệt sau 3 năm với $\alpha = 0,05$.
4. Hãy ước lượng sự khác biệt về tỉ lệ người tiêu dùng thị trường TP Huế có thu nhập trên 8 triệu hiện nay so với 3 năm trước với độ tin cậy 95%.
5. Hãy kiểm định nhận định cho rằng phương sai thu nhập của người tiêu dùng ở thị trường TP Huế đã tăng lên sau 3 năm với $\alpha = 0,01$.

Bài 40. Vị quản đốc cho rằng tỉ lệ thành phẩm của máy A là đã lớn hơn máy B. Để kiểm tra người ta sản xuất thử 1000 sản phẩm trên máy A và 1500 sản phẩm trên máy B. Kết quả cho thấy rằng cả 2 máy đều có 3 sản phẩm hỏng.

1. Hãy kiểm định nhận định trên với $\alpha = 0,01$.
2. Hãy ước lượng sự khác biệt về tỉ lệ thành phẩm của máy A so với máy B với độ tin cậy 95%.

Bài 41. Một doanh nghiệp cho rằng việc nghỉ ngơi giữa giờ sẽ giúp làm tăng tỉ lệ sản phẩm loại đặc biệt. Người ta tiến hành kiển tra trên 50 công nhân được chọn ngẫu nhiên để sản xuất thử mỗi người 100 sản phẩm không nghỉ giữa giờ và 100 sản phẩm có nghỉ giữa giờ. Kết quả như sau:

- Không nghỉ ngơi giữa giờ: Tỉ lệ sản phẩm loại đặc biệt là 5%.
- Có nghỉ ngơi giữa giờ: Tỉ lệ sản phẩm loại đặc biệt là 6%.

1. Hãy kiểm định nhận định trên với $\alpha = 0,01$.
2. Hãy ước lượng sự khác biệt về sản phẩm loại đặc biệt của nghỉ ngơi giữa giờ và không nghỉ ngơi giữa giờ với độ tin cậy 95%.

Bài 42. Cũng công nhân ở Bài 34, ba tháng trước đó, anh ta cũng đã chọn ngẫu nhiên 25 sản phẩm và đo được trọng lượng như sau:

Trọng lượng (kg)	Số sản phẩm
10,1	5
10,2	8
10,3	10
10,4	2

Hãy kiểm định giả thiết cho rằng độ đồng đều của sản phẩm đã giảm đáng kể so với 3 tháng trước, $\alpha = 0,01$. Biết rằng trọng lượng sản phẩm phân phối chuẩn.

Bài 43. Với dữ liệu Bài 15, hãy kiểm định giả thuyết có mối liên hệ tương quan thuận giữa tuổi nghề và năng suất bằng phương pháp kiểm định Pearson với $\alpha = 0,01$.

Bài 44. Với dữ liệu Bài 16, hãy kiểm định giả thuyết cho rằng có liên hệ tương quan giữa chi phí quảng cáo và tỉ suất lợi nhuận trên vốn bằng phương pháp kiểm định Pearson với $\alpha = 0,01$.

Chương 6

Bài 45. Có điểm đánh giá về sự ưa thích của 10 khách hàng được chọn ngẫu nhiên đối với 2 loại sản phẩm A và B trên thang điểm 3 như sau:

Hãy kiểm định giả thuyết cho rằng sản phẩm A được ưa thích bằng hoặc hơn sản phẩm B với $\alpha = 0,05$.

Khách hàng	Điểm sản phẩm A	Điểm sản phẩm B
1	1	2
2	3	2
3	2	1
4	2	1
5	3	2
6	1	2
7	3	2
8	2	1
9	3	1
10	1	2

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Bài 46. Để so sánh sự hài lòng của công nhân đối với cách trả lương mới so với cách trả lương cũ, người ta chọn ngẫu nhiên 8 công nhân và yêu cầu họ cho điểm trên thang điểm 100 đối với hai cách trả lương mới và cũ. Kết quả thu được như sau:

Công nhân	Điểm cho cách trả lương	
	Cũ	Mới
1	53	82
2	62	76
3	44	45
4	37	27
5	72	64
6	36	78
7	69	72
8	58	64

Hãy kiểm định nhận định cho rằng cách trả lương cũ ít được hài lòng hơn cách trả lương mới với $\alpha = 0,05$.

Bài 47. Để so sánh sự hài lòng của công nhân đối với cách trả lương mới so với cách trả lương cũ, người ta chọn 6 công nhân ngẫu nhiên rồi yêu cầu họ cho điểm trên thang điểm 100 đối với cách trả lương mới và chọn 6 công nhân ngẫu nhiên khác rồi yêu cầu họ cho điểm trên thang điểm 100 đối với cách trả lương cũ. Kết quả được sắp xếp theo số điểm như sau:

Điểm cách trả lương cũ	Điểm cách trả lương mới
31	27
33	44
36	45
37	64
44	66
53	67

Hãy kiểm định nhận định cho rằng phương pháp trả lương cũ ít được hài lòng hơn cách trả lương mới với $\alpha = 0,05$.

Bài 48. Để so sánh hiệu quả của 2 phương pháp quảng cáo trước khi đưa vào sử dụng, 14 người được chọn ngẫu nhiên. 7 người được cho xem phương pháp quảng cáo 1. 7 người kia được cho xem phương pháp quảng cáo 2. Sau đó, họ được yêu cầu cho điểm trên thang điểm 100 về mức độ thu hút. Kết quả được sắp xếp theo điểm như sau:

Người được chọn	Điểm phương pháp 1	***Người được chọn***	Điểm phương pháp 2
1	55	***8***	52
2	58	***9***	70
3	59	***10***	73
4	60	***11***	75
5	64	***12***	78
6	65	***13***	85
7	68	***14***	85

Hãy kiểm định cho rằng phương pháp quảng cáo 1 không tốt bằng phương pháp quảng cáo 2 với $\alpha = 0,05$.

Bài 49. Với dữ liệu Bài 15, hãy kiểm định giả thuyết cho rằng có mối liên hệ tương quan thuận giữa tuổi nghề và năng suất bằng phương pháp kiểm định Spearman với $\alpha = 0,01$.

Bài 50. Với dữ liệu Bài 16, hãy kiểm định giả thuyết cho rằng có mối liên hệ tương quan giữa chi phí quảng cáo và tỉ suất lợi nhuận trên vốn bằng phương pháp Spearman với $\alpha = 0,01$.

Bài 51. Có tài liệu về 140 doanh nghiệp được chọn ngẫu nhiên ở một thành phố và được phân tổ kết hợp theo quy mô và tỉ suất lợi nhuận trên vốn như sau:

Quy mô	Tỉ suất lợi nhuận trên vốn (%)			
	5-10	10-15	15-20	Cộng
Nhỏ	20	60	6	86
Vừa	5	30	19	54
Cộng	25	90	25	140

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Hãy kiểm định giả thuyết về liên hệ thuận giữa quy mô và tỉ suất lợi nhuận trên vốn với $\alpha = 0,05$ bằng phương pháp Kendal Tau A.

Bài 52. Có tài liệu về 110 doanh nghiệp được chọn ngẫu nhiên ở một thành phố và được phân tổ kết hợp theo tỉ suất lợi nhuận và xếp loại tín dụng như sau:

Tỉ suất lợi nhuận (%)	Xếp loại tín dụng		
	Kém	Trung bình	Khá
Dưới 10	18	10	7
10-14	8	15	12
Trên 14	5	9	26

Hãy kiểm định giả thuyết về liên hệ thuận giữa tỉ suất lợi nhuận và mức lương giám đốc với $\alpha = 0,05$ bằng kiểm định Gammar.

Bài 53. Để nghiên cứu mối liên hệ giữa giới tính và loại bao bì của một loại sản phẩm, một doanh nghiệp chọn ngẫu nhiên 100 khách hàng gồm 60 nam và 40 nữ để điều tra xem họ thích chọn loại bao bì nào. Kết quả điều tra được phân tổ như trong bảng trang sau.

Hãy kiểm định tính độc lập giữa giới tính và loại bao bì lựa chọn với $\alpha = 0,05$.

Giới tính	Loại bao bì lựa chọn		
	A	B	C
Nam	33	20	7
Nữ	8	10	22

Bài 54. Có tài liệu về 140 doanh nghiệp được chọn ngẫu nhiên ở một thành phố và được phân tổ kết hợp theo loại hình kinh doanh và tỉ suất lợi nhuận trên vốn như sau:

Loại hình kinh doanh	Tỉ suất lợi nhuận trên vốn (%)		
	5-10	10-15	15-20
Sản suất	20	60	6
Dịch vụ	5	30	19

Hãy kiểm định tính độc lập giữa loại hình kinh doanh và tỉ suất lợi nhuận trên vốn với $\alpha = 0,05$.

Bài 55. Có nhận định cho rằng phân phối thu nhập của khách hàng siêu thị ở bài 4 như sau:

Thu nhập (tr.đồng)	Dưới 5	5 - 10	10 - 30	30 - 60	Trên 60
Tần suất (%)	15	25	40	15	5

Hãy kiểm định nhận định trên với $\alpha = 0,05$.

Bài 56. Một công ty muốn kiểm tra hiểu biết của khách hàng về một tính năng sản phẩm mà công ty đưa vào sản phẩm đã được một thời gian. 7 câu hỏi về tính năng, mỗi câu có sẵn 5 lựa chọn trả lời trong đó chỉ có một lựa chọn trả lời đúng, đã được đặt ra cho 150 khách hàng được chọn ngẫu nhiên. Kết quả thu được như sau:

Số câu trả lời đúng	0	1	2	3	4	5	6
Số khách trả lời đúng	12	15	28	34	26	22	13

Yêu cầu:

1. Lập phân phối xác suất cho số câu trả lời đúng của một khách hàng nếu khách hàng này hoàn toàn không biết gì về tính năng công ty đang khảo sát mà chỉ trả lời ngẫu nhiên. *(gợi ý: xem bài tập 19)*
2. Kiểm định giả thuyết cho rằng các khách hàng hoàn toàn không biết gì về tính năng mà công ty đang khảo sát bằng phương pháp khi bình phương với sai lầm loại 1 là 5%.

Bài 57 . Một công ty muốn nghiên cứu ảnh hưởng của bao bì đối với một loại sản phẩm. Công ty chọn ngẫu nhiên 200 gia đình và cho dùng thử 4 sản phẩm cùng loại nhưng được đựng trong 4 loại bao bì khác nhau (công ty nói với 200 gia đình này đó là 4 loại sản phẩm khác nhau). Một tháng sau, khi được hỏi gia đình sẽ chọn loại sản phẩm nào, kết quả thu được như sau:

Loại bao bì	A	B	C	D
Số gia đình chọn	42	58	65	35

Hãy kiểm định giả thuyết cho rằng bao bì không ảnh hưởng đến sự lựa chọn của các khách hàng với sai lầm loại 1 là 5% bằng phương pháp Khi bình phương. *(gợi ý: bao bì không ảnh hưởng đến sự lựa chọn thì xác suất lựa chọn mỗi loại bao bì là như nhau)*

Bài 58. Với dữ liệu Bài 86. yêu cầu:

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

1. Kiểm định giả thuyết về phân phối chuẩn của quảng cáo A bằng kiểm định Kolmogorov-smirnov với α =0,05.
2. Kiểm định giả thuyết về phân phối chuẩn của quảng cáo B bằng kiểm định Kolmogorov-smirnov với sai lầm loại 1 là 0,01.
3. Kiểm định giả thuyết về phân phối chuẩn của quảng cáo C bằng kiểm định Kolmogorov-smirnov với sai lầm loại 1 là 0,05 .

Chương 7

Bài 59. Có dữ liệu thu thập về 20 khách hàng được chọn ngẫu nhiên của một siêu thị như trong bảng sau.

Tuổi	Giới tính	Nghề nghiệp	Thu nhập (tr. Đồng)	Giá trị hóa đơn (tr.đ)	Số lần mua hàng
27	Nam	A	2,5	0,8	2
32	Nữ	C	1,8	0,5	3
19	Nam	B	3,4	1,4	4
28	Nam	A	4,5	2,2	5
37	Nữ	B	5,6	3,0	4
45	Nữ	C	8,2	4,5	10
38	Nam	B	4,8	2,5	7
42	Nam	A	5,0	2,8	8
60	Nữ	C	2,2	1,0	1
56	Nam	B	3,5	2,0	4
48	Nữ	A	2,9	1,5	2
52	Nam	B	8,5	4,0	7
47	Nam	A	9,3	5,0	14
53	Nam	C	3,0	1,6	3
34	Nữ	A	5,2	2,6	9
22	Nữ	B	3,7	2,4	4
50	Nữ	A	4,2	2,9	6
44	Nữ	C	2,7	1,2	2
34	Nam	B	4,8	3,2	6
47	Nam	A	7,4	3,6	11

Trong đó: A: Chính khách và doanh nhân; B: Văn nghệ sĩ và nhà khoa học; C: Nghề nhiệp khác
1. Xác định hàm hồi quy tuyến tính mô tả mối liên hệ giữa Giá trị hóa đơn và Thu nhập hàng tháng. Giải thích ý nghĩa các hệ số hồi qui thu được.
2. Xác định R^2 và giải thích ý nghĩa.
3. Kiểm định t về sự phù hợp của mô hình. Biết rằng phương sai mẫu của thu nhập (trên 20 khách hàng) là 4,75.
4. Kiểm định F về sự phù hợp của mô hình.

Bài 60. Có dữ liệu về tuổi nghề và năng suất lao động của 5 công nhân được chọn ngẫu nhiên ở một doanh nghiệp như sau:

Công nhân	Tuổi nghề	Năng suất (SP/giờ)
1	2	2
2	1	3
3	5	6
4	3	4
5	6	8

1. Xác định hàm hồi quy tuyến tính mô tả mối liên hệ giữa tuổi nghề và năng suất. Giải thích ý nghĩa các hệ số hồi qui thu được.
2. Xác định R^2 và giải thích ý nghĩa.
3. Kiểm định t về sự phù hợp của mô hình. Biết rằng $\sum(X_i - \bar{X})^2 = 17{,}2$.
4. Kiểm định F về sự phù hợp của mô hình.

Bài 61. Với dữ liệu bài 59, yêu cầu:
1. Xác định hàm hồi quy loga mô tả mối liên hệ giữa Giá trị hóa đơn và Thu nhập hàng tháng.
2. Xác định R^2 và giải thích ý nghĩa.
3. Kiểm định t về sự phù hợp của mô hình.
4. Kiểm định F về sự phù hợp của mô hình.

Bài 62. Với dữ liệu bài 59, yêu cầu:
1. Xác định hàm hồi quy Hypecbon mô tả mối liên hệ giữa Số lần mua hàng và Thu nhập hàng tháng.
2. Xác định R^2 và giải thích ý nghĩa.
3. Kiểm định t về sự phù hợp của mô hình.
4. Kiểm định F về sự phù hợp của mô hình.

Bài 63. Với dữ liệu bài 59, yêu cầu:
1. Viết các ma trận dữ liệu của mô hình hồi quy tuyến tính bội mô tả mối liên hệ giữa Giá trị hóa đơn, Thu nhập hàng tháng, Tuổi.
2. Viết mô hình tuyến tính bội mô tả mối liên hệ giữa Giá trị hóa đơn, Thu nhập hàng tháng, Tuổi và giải thích ý nghĩa các hệ số hồi quy.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

3. Xác định R^2 và giải thích ý nghĩa.
4. Kiểm định t về sự phù hợp của mô hình.
5. Kiểm định F về sự phù hợp của mô hình.

Bài 64. Với dữ liệu bài 59, yêu cầu:
1. Lượng hóa tiêu thức giới tính bằng biến giả.
2. Viết các ma trận dữ liệu của mô hình hồi quy tuyến tính bội mô tả mối liên hệ giữa Số lần mua hàng ở siêu thị, Thu nhập hàng tháng, Tuổi, Giới tính.
3. Viết mô hình tuyến tính bội mô tả mối liên hệ giữa Số lần mua hàng ở siêu thị, Thu nhập hàng tháng, Tuổi, Giới tính và giải thích ý nghĩa các hệ số hồi quy.
4. Xác định R^2 và giải thích ý nghĩa.
5. Kiểm định t về sự phù hợp của mô hình.
6. Kiểm định F về sự phù hợp của mô hình.

Bài 65. Có dữ liệu thu thập được trong 19 năm của một doanh nghiệp như sau.

Năm	Lương giám đốc (tr.đ)	Lương bình quân một công nhân viên (tr.đ)	Doanh số (tỉ đồng)	Tỉ suất lợi nhuận trên vốn (%)
1996	3,4	1,3	5,9	13,4
1997	3,5	1,3	6,5	11,6
1998	3,8	1,5	5,7	12,8
1999	4,2	1,9	7,2	11,9
2000	4,7	2,0	6,7	12,5
2001	5,0	2,0	8,0	13,2
2002	5,3	2,2	8,4	13,5
2003	5,5	2,3	7,6	12,5
2004	5,5	2,3	8,8	12,9
2005	5,9	2,4	9,0	14,0
2006	6,0	2,7	8,2	14,4
2007	6,0	2,7	9,2	13,1
2008	6,5	3,0	9,9	13,5
2009	6,5	3,0	8,5	12,7
2010	7,0	3,4	9,4	14,8
2011	7,5	3,4	9,0	13,5
2012	8,5	3,7	10,4	14,6
2013	10,0	4,0	11,5	15,4
2014	9,8	4,1	11,3	15,6

Kết quả hồi quy trên SPSS mối liên hệ tuyến tính giữa Tỉ suất lợi nhuận trên vốn và Thời gian như sau:

Coefficients

	Unstandardized Coefficients		t	Sig.
	b_j	Std. Error		
(Constant)	-284	62.2	-4.568	.000
Thoi gian	.148	.031	4.785	.000

Model Summary

R	R Square	Adjusted R Square	Std. Error	Durbin-Watson
.758	.574	.549	.741	2.057

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	12.556	1	12.56	22.89	.000
Residual	9.325	17	.549		
Total	21.881	18			

1. Mô hình có hiện tượng tự tương quan hay không ?
2. Mô hình có bảo đảm ý nghĩa ở mức 1% hay không?
3. Mức độ giải thích của mô hình là bao nhiêu?
4. Viết mô hình hồi quy mẫu, giải thích ý nghĩa các hệ số hồi quy.

Bài 66. Với dữ liệu ở Bài 65. có kết quả hồi quy trên SPSS mối liên hệ tuyến tính giữa Tỉ suất lợi nhuận trên vốn, Thời gian, Lương giám đốc, Lương bình quân một công nhân viên, Doanh số như sau:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error	Durbin-Watson
4	.82	.67	.66	.65	2.41

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
4	Regression	14.	1	14.77	35.29	.000
	Residual	7.11	17	.42		
	Total	21.88	18			

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Coefficients

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics	
		b_j	Std. Error			Tol	VIF
1	(Constant)	279	416	.67	.513	-	-
	Thoi gian	-.135	.210	-.65	.529	.019	51.6
	Luong GD	.544	.500	1.1	.295	.030	33.6
	LuongCN	.580	1.87	.31	.761	.010	95.5
	DoanhSo	.074	.309	.24	.813	.104	9.6
2	(Constant)	252	388	.65	.526	-	-
	Thoi gian	-.121	.195	-.62	.543	.021	47.6
	Luong GD	.599	.430	1.39	.184	.038	26.5
	LuongCN	.501	1.79	.28	.783	.011	92. 6
3	(Constant)	158	193	.82	.423	-	-
	Thoi gian	-.074	.097	-.77	.454	.080	12.5
	Luong GD	.687	.287	2.40	.029	.080	12.5
4	(Constant)	10.6	.506	21.0	.000	-	-
	Luong GD	.476	.080	5.94	.000	1	1

1. Có hiện tượng đa cộng tuyến ở mô hình nào? Vì sao?
2. SPSS lựa chọn biến loại khỏi mô hình dựa vào căn cứ nào?
3. Mô hình 4 có hiện tượng tự tương quan hay không ?
4. Mô hình 4 có bảo đảm ý nghĩa ở mức 1% hay không?
5. Mức độ giải thích của mô hình 4 là bao nhiêu?
6. Viết hàm hồi quy mẫu, giải thích ý nghĩa các hệ số hồi quy.

Chương 8

Bài 67. Có dữ liệu về một doanh nghiệp như sau.

	Tháng 1	Tháng 2	Tháng 3	Tháng 4
- Sản lượng (tấn)	300	500	400	600
- Số công nhân đầu tháng	200	205	203	207

1. Xác định sản lượng trung bình một tháng trong quí I.
2. Xác định số công nhân trung bình trong quí I.
3. Xác định số công nhân trung bình trong từng tháng.

Bài 68. Có số lượng xe đổ xăng đếm được ở một cây xăng tại một số thời điểm như sau:

- Lúc 5h có 5 xe
- Lúc 5h30 có 8 xe
- Lúc 6h30 có 12 xe
- Lúc 7h10 có 6 xe
- Lúc 7h30 có 3 xe
- Lúc 8h có 1 xe

1. Vẽ biểu đồ mô tả biến động số lượng xe đổ xăng từ 5h đến 8h.
2. Hãy xác định số lượng xe đổ xăng trung bình tại cây xăng từ 5h đến 8h.

Bài 69. Có tài liệu về số lượng công nhân của một doanh nghiệp như sau:

- Ngày 1/2 có 120 công nhân;
- Ngày 15/2 tuyển thêm 5 công nhân;
- Ngày 20/2 cho thôi việc 2 công nhân;
- Ngày 25/2 tuyển thêm 4 công nhân;
- Ngày 3/3 cho thôi việc 1 công nhân.

1. Vẽ biểu đồ mô tả biến động số lượng công nhân của doanh nghiệp trong tháng 2.
2. Hãy xác định số công nhân trung bình trong tháng 2.

Bài 70. Số sinh viên trong 3 tiết học từ 7h đến 9h40 của một lớp học như sau:

- Lúc 7h có 40 sinh viên;
- Lúc 7h20 có 4 sinh viên đi học trễ vào lớp;
- Lúc 8h10 có 3 sinh viên xin nghĩ học đi khám sức khỏe;
- Lúc 8h30 có 2 sinh viên trốn học ra về;
- Lúc 9h20 có 1 sinh viên xin về sớm để đi thi tại trung tâm ngoại ngữ.

1. Vẽ biểu đồ mô tả biến động số sinh viên trong 3 tiết học từ 7h đến 9h40.
2. Hãy xác định số sinh viên trung bình trong 3 tiết học từ 7h đến 9h40.

Bài 71. Có dữ liệu về dự trữ hàng hóa của một doanh nghiệp như sau:

Ngày	1/1	1/2	1/3	1/4	1/5	1/6	1/7
Mức dự trữ (tỉ đồng)	1,2	2,8	2,0	2,7	1,2	1,8	2,5

1. Vẽ biểu đồ mô tả biến động dự trữ hàng hóa của doanh nghiệp 6 tháng đầu năm.
2. Hãy xác định mức dự trữ trung bình trong 6 tháng đầu năm của doanh nghiệp.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Bài 72. Có dữ liệu về vốn kinh doanh của một doanh nghiệp như sau:

Ngày	1/1	12/1	14/2	25/2	4/3	26/3	1/4
Vốn (tỉ đồng)	1,4	2,2	2,8	1,9	2,5	1,5	2,8

1. Vẽ biểu đồ mô tả biến động vốn kinh doanh của doanh nghiệp trong quí I.
2. Hãy xác định vốn kinh doanh trung bình trong quí I của doanh nghiệp.

Bài 73. Có dãy số thời gian về sản lượng của một doanh nghiệp như sau:

Năm	2012	2013	2014	2015	2016	2017	2018
Doanh thu thuần (tỉ đ)	4	7	6	9	10	13	12

1. Xác định doanh thu thuần trung bình một năm
2. Tính lượng tăng (giảm) tuyệt đối trung bình một năm về doanh thu thuần;
3. Hãy xác định Tốc độ phát triển doanh thu thuần trung bình một năm.
4. Hãy làm nhẵn dãy số sản lượng bằng phương pháp trung bình trượt với bước trượt là 2 mức độ, mô tả các dãy số trên đồ thị và cho nhận xét;
5. Mô tả xu hướng phát triển cơ bản của doanh thu thuần từ 2012 đến 2018 bằng hàm hồi qui tuyến tính.
6. Hãy dự đoán doanh thu thuần cho các năm 2019, 2020 theo tất cả các phương pháp đã học thích hợp. Chọn $\alpha = 0,7$; $\beta = 0,6$.

Bài 74. Có dãy số thời gian về lợi nhuận của một doanh nghiệp như sau:

Năm	2013	2014	2015	2016	2017	2018
Tốc độ phát triển lợi nhuận so với năm 2012 (%)	105	110	114	115	112	116

1. Hãy xác định tốc độ phát triển lợi nhuận trung bình một năm của giai đoạn 2012 đến 2016.
2. Biết rằng lợi nhuận năm 2016 là 12 tỉ đồng. Hãy xác định lợi nhuận 2012 và 2018.

Bài 75. Có thông tin về sản lượng của một doanh nghiệp như sau: trong 3 năm từ 2002 đến 2005 tốc độ phát triển sản lượng trung bình là 102% mỗi năm; trong 2 năm từ 2005 đến 2007 tốc độ phát triển sản lượng trung bình là 105% mỗi năm; trong 5 năm từ 2007 đến 2012 tốc độ phát triển sản lượng trung bình là 99% mỗi năm; trong 4 năm từ 2012 đến 2016 tốc độ phát triển sản lượng trung bình là 107% mỗi năm.

1. Hãy xác định tốc độ tăng sản lượng từ 2002 đến 2005.
2. Hãy xác định tốc độ tăng sản lượng trung bình một năm của giai đoạn từ 2002 đến 2016.
3. Hãy xác định sản lượng năm 2002 biết rằng sản lượng năm 2016 của doanh nghiệp là 20 tỉ đồng.

Bài 76. Doanh số bán của một doanh nghiệp giảm trung bình 3% hàng năm trong 5 năm từ 1995 đến 2000, tăng trung bình 6% hàng năm trong 4 năm từ 2000 đến 2004, tăng trung bình 4% hàng năm trong 3 năm từ 2004 đến 2007.
1. Hãy xác định tốc độ tăng doanh số bán cho giai đoạn từ 1995 đến 2000.
2. Hãy xác định tốc độ tăng trung bình hàng năm cho giai đoạn từ 1995 đến 2007.

Bài 77. Có dữ liệu như sau về một doanh nghiệp:

Năm	2011	2012	2013	2014	2015
Tốc độ tăng giá bán liên hoàn (%)	4	6	5	2	3
Tốc độ tăng lượng bán liên hoàn (%)	10	12	8	6	14

1. Lập dãy số Tốc độ phát triển giá bán liên hoàn.
2. Xác định tốc độ tăng giá bán trung bình một năm giai đoạn 2010-2015.
3. Xác định tốc độ phát triển lượng bán năm 2015 so với 2012.
4. Biết rằng lượng bán năm 2015 là 10.000 sản phẩm. Hãy xác định lượng bán năm 2014 và lượng bán 2013.

Bài 78. Có dữ liệu như sau về một doanh nghiệp:

Năm	2014	2015	2016	2017	2018
Tốc độ tăng giá thành liên hoàn	0,05	0,04	0,06	-0,03	-0,02
Tốc độ tăng sản lượng liên hoàn	0,12	0,10	0,07	0,05	0,11

1. Lập dãy số Tốc độ phát triển giá thành liên hoàn.
2. Xác định tốc độ tăng giá thành 2018 so với 2013.
3. Xác định tốc độ tăng sản lượng trung bình một năm giai đoạn 2015-2018.
4. Biết rằng giá thành năm 2016 là 2 triệu/sản phẩm. Hãy xác định giá thành năm 2017, 2018.

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Bài 79. Có dữ liệu về doanh số mua vào một loại hàng thuỷ sản của một doanh nghiệp chế biến như sau:

Quí	Doanh số mua (tỉ đồng)		
	2016	2017	2018
I	3,0	3,4	3,5
II	5,6	6,3	7,0
III	2,4	3,0	3,1
IV	1,0	1,3	1,4
Tổng cộng	12	14	15

1. Hãy mô tả xu hướng phát triển cơ bản của doanh số mua (theo quí) bằng số trung bình trượt, mô tả các dãy số trên đồ thị và cho nhận xét.
2. Hãy xác định chỉ số thời vụ giản đơn và biểu diễn trên đồ thị.
3. Hãy xác định chỉ số thời vụ và biểu diễn trên đồ thị.
4. Hãy phân tích các thành phần của dãy số thời gian theo mô hình nhân.
5. Hãy dự đoán tổng doanh số mua vào cho các năm 2019 theo tất cả các phương pháp đã học. Chọn $\alpha = 0,8$; $\beta = 0,7$.

Bài 80. Có dữ liệu về doanh số bán của một doanh nghiệp sản xuất vật liệu trang trí nội thất như sau:

Quí	Doanh số (tỉ đồng)			
	2015	2016	2017	2018
I	0,9	1,0	1,2	1,8
II	2,6	3,0	3,0	4,1
III	1,8	2,0	2,6	3,6
IV	4,7	5,0	5,2	6,5
Tổng cộng	10	11	12	16

1. Hãy xác định chỉ số thời vụ giản đơn và biểu diễn trên đồ thị;
2. Hãy xác định chỉ số thời vụ và biểu diễn trên đồ thị;
3. Hãy phân tích các thành phần của dãy số thời gian theo mô hình nhân.
4. Hãy dự đoán tổng doanh số bán cho các năm 2019, 2020 theo tất cả các phương pháp đã học. Chọn $\alpha = 0,6$; $\beta = 0,8$.

Bài 81. Có tài liệu về một cửa hàng như sau:

Loại hàng	ĐVT	Đơn giá (1000đ)		Lượng bán	
		Tháng 1	Tháng 2	Tháng 1	Tháng 2
- Gạo	Kg	3,5	3,8	4500	6000
- Thịt bò	Kg	80,0	90,0	1600	900
- Dầu ăn	l	10,0	11,0	300	500

1. Hãy xác định các chỉ số giá cá thể
2. Hãy xác định các chỉ số lượng bán cá thể
3. Hãy xác định chỉ số giá tổng hợp Paasche;
4. Hãy xác định chỉ số lượng bán tổng hợp.

Bài 82. Theo phương pháp định giá hàng tồn kho nhập sau-xuất trước (LIFO), một chỉ số giá tổng hợp cho hàng tồn kho phải được xác định cho mục đích tính thuế. Trong đó, trọng số khối lượng là mức tồn kho cuối năm, đơn giá đầu năm làm giá kỳ gốc. Sau đây là dữ liệu về một số loại sản phẩm tồn kho được chọn đại diện cho một kho hàng.

Loại sản phẩm	Đơn giá (tr.đ)		Tồn kho cuối năm
	Đầu năm	Cuối năm	
A	0,15	0,19	500
B	1,60	1,80	50
C	4,50	4,20	100
D	12,00	13,20	40

Hãy tính chỉ số giá tổng hợp có trọng số dùng cho mục đích định giá hàng tồn kho LIFO của kho hàng trên.

Bài 83. Có tài liệu về một cửa hàng như sau.

Loại hàng	Doanh số bán năm 2014 (tr.đ)	Doanh số bán năm 2015 (tr.đ)	Chỉ số cá thể	
			i_p (%)	i_q (%)
- Vải thun	500	700	90	130
- Áo trẻ em	400	400	110	120
- Quần Jean	800	1200	115	110
- Len	450	600	95	140

1. Biết rằng lượng bán vải thun năm 2015 là 80 nghìn mét. Hãy xác định lượng bán vải thun năm 2014.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

2. Hãy xác định Chỉ số giá tổng hợp Laspeyres.
3. Hãy xác định Chỉ số lượng bán tổng hợp.

Bài 84. Có dữ liệu như sau về một doanh nghiệp:

Loại hàng	Doanh số tiêu thụ (triệu đồng)		Tốc độ tăng giá bán (%)
	Tháng 1	Tháng 2	
- Áo	600	900	+6
- Quần	1500	1400	-1

1. Biết rằng đơn giá bán của quần trong tháng 1 là 125 nghìn đồng/cái. Hãy xác định đơn giá bán của quần trong tháng 2.
2. Hãy tính chỉ số giá tổng hợp Laspeyres của doanh nghiệp.
3. Hãy tính chỉ số lượng bán tổng hợp của doanh nghiệp.

Bài 85. Công ty hóa chất Mitchell sản xuất một hóa chất công nghiệp đặc biệt. Đó là một hỗn hợp của 3 thành phần. Dữ liệu về đơn giá mua và tỉ lệ hỗn hợp 3 nguyên liệu như sau.

Nguyên liệu	Đơn giá mua (tr.đ)		Khối lượng cho 100 kg sản phẩm
	Đầu năm	Cuối năm	
A	2,50	3,95	25
B	8,75	9,90	15
C	0,99	0,95	60

1. Hãy tính chỉ số giá cá thể cho từng loại nguyên liệu.
2. Tính chỉ số giá tổng hợp từ các chỉ số giá cá thể cho các nguyên liệu thô dùng trong sản phẩm. Giải thích ý nghĩa của chỉ số.

Bài 86. Một công ty vận tải nhận vận chuyển bốn loại hàng cho một nhà phân phối. Dữ liệu về số chuyến và đơn giá được báo cáo như bảng sau.

Loại hàng	Số chuyến		Đơn giá mỗi chuyến năm 1996 (tỉ đ)
	2010	2016	
A	120	95	1,2
B	86	75	1,8
C	35	50	2,0
D	60	70	1,5

Hãy tính chỉ số khối lượng vận chuyển tổng hợp của công ty năm 2016 so với năm 2010. Nhận xét về biến động khối lượng vận chuyển của công ty.

Chương 9

Bài 87. Một nhà sản xuất muốn so sánh hiệu quả tác động của 3 cách quảng cáo khác nhau đến sự hứng thú của khách hàng. Một dụng cụ đo đồng tử được sử dụng. Nhà sản xuất chọn ngẫu nhiên 18 người để thực nghiệm. Trong đó 6 người được cho xem quảng cáo A (hứa hẹn nguồn lợi trực tiếp), 6 người được cho xem quảng cáo B (gây sự tò mò), và 6 người còn lại được cho xem quảng cáo C có tính chất so sánh). Kết quả đo được trên dụng cụ đo đồng tử như sau:

Quảng cáo A	Quảng cáo B	Quảng cáo C
8	3	4
8	4	7
9	2	6
7	1	3
8	3	5
9	2	3

1. Dùng phương pháp kiểm định hạng Kruskal-Wallis để kiểm định giả thiết về sự bằng nhau của hiệu quả của 3 loại quảng cáo với $\alpha = 0,05$.
2. Thiết kế trên thuộc loại gì ? Vì sao ?
3. Hãy tiến hành một phân tích phương sai để so sánh hiệu quả của 3 loại quảng cáo với $\alpha = 0,05$.
4. So sánh bội giữa các loại quảng cáo với $\alpha = 0,05$. (nếu cần).

Bài 88. Người ta muốn kiểm tra sự ưa thích của khách hàng đối với 3 loại sản phẩm. Một nhóm 21 khách hàng được chọn ngẫu nhiên để xem sản phẩm và cho điểm từ 0 đến 10 cho mỗi loại sản phẩm. Kết quả thu được như sau:

Loại sản phẩm	Điểm số						
1	3	2	3	5	4	1	3
2	7	6	9	9	6	8	7
3	7	5	4	3	6	4	5

1. Dùng phương pháp kiểm định hạng Kruskal-Wallis để kiểm định giả thiết về sự ưa thích như nhau đối với 3 loại sản phẩm với $\alpha = 0,05$.
2. Thiết kế trên thuộc loại gì ? Vì sao ?
3. Hãy tiến hành một phân tích phương sai về sự ưa thích như nhau của khách hàng đối với 3 loại sản phẩm với $\alpha = 0,05$.
4. So sánh bội sự ưa thích của khách hàng đối với mỗi loại sản phẩm với $\alpha = 0,05$. (nếu cần).

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Bài 89. Một nhà sản xuất muốn so sánh tác dụng của 3 loại quảng cáo: Quảng cáo A (thiên về giá cả), quảng cáo B (thiên về chất lượng), quảng cáo C (thiên về tiện lợi) cho một loại thức ăn đông lạnh. Một thực nghiệm được tiến hành trên 4 thành phố. Tại mỗi thành phố, người ta chọn 3 cửa hàng, mỗi cửa hàng cho một loại quảng cáo một cách ngẫu nhiên. Doanh số bán tăng lên trong một tháng sau khi quảng cáo tính bằng triệu đồng thu được như sau:

Thành phố	Quảng cáo A	Quảng cáo B	Quảng cáo C
1	9	2	3
2	9	1	5
3	10	4	4
4	8	5	2

1. Thiết kế trên thuộc loại gì ? Vì sao ?
2. Hãy tiến hành một phân tích phương sai để so sánh hiệu quả của 3 loại quảng cáo với $\alpha = 0,05$.
3. So sánh bội giữa các loại quảng cáo với $\alpha = 0,05$. (nếu cần).

Bài 90. Để so sánh hiệu quả của 5 loại phân bón A, B, C, D, E người ta tiến hành thử nghiệm ở 5 nông trại. Mỗi nông trại được chọn ra 5 lô đất để trồng ngô và bón phân tương ứng với sơ đồ thiết kế như sau: (năng suất tính bằng tạ/ha)

Nông trại	Lô đất				
	1	2	3	4	5
1	D 10	E 6	A 2	C 7	B 8
2	E 8	B 6	C 3	A 2	D 9
3	A 3	C 4	B 6	D 11	E 8
4	C 3	D 9	E 8	B 6	A 2
5	B 7	A 1	D 12	E 9	C 3

1. Thiết kế trên thuộc loại gì ? Vì sao ?
2. Hãy tiến hành một phân tích phương sai để so sánh hiệu quả của 4 loại phân với $\alpha = 0,05$.
3. So sánh bội giữa 5 nông trại với $\alpha = 0,05$. (nếu cần).
4. So sánh bội giữa 5 loại phân với $\alpha = 0,05$. (nếu cần).

Bài 91. Một nhà kinh doanh thương mại muốn so sánh hiệu quả của 3 cách thức chiếu sáng cửa hàng:
- Cách A: Ánh sáng màu dịu
- Cách B: Ánh sáng trắng rực rỡ
- Cách C: Ánh sáng màu rực rỡ

Nhà kinh doanh này lo ngại việc thử nghiệm liên tiếp thì kết quả cách chiếu sáng trước có thể ảnh hưởng đến kết quả cách chiếu sáng sau; các cửa hàng khác nhau cũng có thể có tác dụng khác nhau. Thiết kế thực nghiệm và doanh số bán (chục triệu) trong các thời gian thực nghiệm thu được như sau:
1. Thiết kế trên thuộc loại gì ? Vì sao ?
2. Hãy tiến hành một phân tích phương sai để so sánh hiệu quả của 3 cách chiếu sáng với $\alpha = 0,05$.
3. So sánh bội giữa các cách chiếu sáng với $\alpha = 0,05$. (nếu cần).

Cửa hàng số	Thời gian thực nghiệm		
	3/9-16/9	23/9-7/10	14/10-27/10
1	5(A)	9(B)	5 (C)
2	8 (B)	7 (C)	3(A)
3	2 (C)	4(A)	9 (B)

Bài 92. Một nhà quản lý muốn đánh giá ảnh hưởng của âm nhạc đến năng suất làm việc của công nhân. 4 chương trình nhạc A,B,C,D (nhẹ, rock, cổ điển, dân tộc) và không có âm nhạc E được so sánh với nhau. Để tránh ảnh hưởng lẫn nhau, người ta tiến hành thử nghiệm trong 5 tuần liên tiếp theo sơ đồ thiết kế như sau: (năng suất tính bằng sp/giờ trong một ngày làm việc)

Tuần	Thứ 2	Thứ 3	Thứ 4	Thứ 5	Thứ 6
1	A 4	B 4	C 6	D 7	E 2
2	B 8	C7	D 7	E 2	A 1
3	C 8	A 3	E 3	B 4	D 6
4	D 9	E 4	A 2	C 6	B 3
5	E 5	D 8	B 4	A 4	C 5

1. Thiết kế trên thuộc loại gì ? Vì sao ?
2. Hãy tiến hành một phân tích phương sai để so sánh tác dụng của 4 loại nhạc và đối chứng không nhạc với $\alpha = 0,05$.
3. So sánh bội ảnh hưởng của các chương trình âm nhạc với $\alpha = 0,05$. (nếu cần).

Bài 93. Một công ty xăng dầu muốn so sánh hiệu quả của 4 loại xăng A,B,C,D dựa trên số km chạy được cho 1 lít xăng. Các tài xế điều khiển xe và kiểu xe có thể có ảnh hưởng đến kết quả thử nghiệm. Vì vậy, người ta sử dụng thiết kế vuông ngẫu nhiên. Người ta chọn ngẫu nhiên 16 xe với 4 xe mỗi kiểu và 16 thùng xăng với 4 thùng cho 1 loại xăng và 4 tài xế. Sơ đồ thiết kế vuông và kết quả đo được về đoạn đường chạy được (trăm km) cho 10 lít xăng tương ứng được trình bày trên bảng sau:

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Tài xế	Kiểu xe			
	1	2	3	4
1	A 9	B 2	C 7	D 5
2	B 3	C 7	D 6	A 10
3	C 6	D 5	A 8	B 3
4	D 6	A 9	B 2	C 6

Hãy kiểm định về sự giống nhau về hiệu quả của 4 loại xăng với $\alpha = 0,05$.

Bài 94. Một nhà kinh doanh cà phê muốn chọn hai hình thức quảng bá cho sản phẩm của mình:

- Quảng cáo trên đài truyền hình địa phương.
- Hạ giá bán.

Ông ta cho rằng, trên thị trường đầy cạnh tranh, nếu chỉ quảng cáo thôi thì chưa đủ mà phải đồng thời giảm giá để hấp dẫn khách mua hàng. Tuy nhiên, nếu quảng cáo hay giảm giá thái quá có thể có tác tác dụng tiêu cực. Do đó, ông thử nghiệm với 3 mức giảm giá và 3 mức chi phí quảng cáo (tính bằng % trên doanh thu lúc chưa thử nghiệm). Ông chọn ngẫu nhiên 18 thành phố gần giống nhau để thử nghiệm. Mỗi kết hợp mức giảm giá-mức chi phí quảng cáo được tiến hành ở 2 thành phố. Mức tăng lượng bán 1 tháng thử nghiệm ở mỗi thành phố là như sau (tính bằng 100kg):

Mức giảm giá (1000/kg)	Chi phí quảng cáo (%)		
	0,6	1	1,4
3	2	3	4
	1	4	5
4	3	4	8
	4	5	9
5	5	6	6
	4	5	7

1. Thiết kế trên thuộc loại gì ? Vì sao ?
2. Hãy tiến hành một phân tích phương sai để kiểm định sự không tồn tại tương tác giữa 2 yếu tố mức giảm giá và chi phí quảng cáo với $\alpha = 0,05$.
3. So sánh bội giữa các mức chi phí quảng cáo với $\alpha = 0,05$. (nếu cần).
4. So sánh bội giữa các mức giảm giá với $\alpha = 0,05$. (nếu cần).
5. So sánh bội giữa các kết hợp mức giảm giá-mức chi phí quảng cáo với $\alpha = 0,05$. (nếu cần).

Bài 95. Một doanh nghiệp thương mại muốn nghiên cứu ảnh hưởng của 2 nhân tố là chính sách khuyến mại và cách chiếu sáng đến doanh số bán. Chính sách khuyến mại có 2 mức độ là bốc thăm trúng thưởng lớn (A_1), tặng quà theo định lượng mua (A_2). Cách chiếu sáng có 3 mức độ là: Anh sáng màu dịu (B_1), ánh sáng trắng rực rỡ (B_2), ánh sáng màu rực rỡ (B_3). Như vậy có 6 kết hợp thử nghiệm. Mỗi kết hợp được thử nghiệm 4 lần, mỗi lần là một tuần bán hàng cho kết hợp đó. Để tránh ảnh hưởng kéo dài của thử nghiệm kế trước, trước mỗi thử nghiệm là 1 tuần bán hàng theo không khuyến mại và chiếu sáng thông thường. Kết quả của mỗi tuần thử nghiệm là doanh số bán hàng mỗi tuần (tính bằng triệu đồng), thu thập được như sau:

1. Thiết kế trên thuộc loại gì ? Vì sao ?
2. Hãy tiến hành một phân tích phương sai để kiểm định sự không tồn tại tương tác giữa 2 yếu tố chính sách khuyến mại và cách chiếu sáng với $\alpha = 0,05$.
3. So sánh bội giữa các kết hợp chính sách khuyến mại và cách chiếu sáng với $\alpha = 0,05$ (nếu cần).

Chính sách khuyến mại	Cách chiếu sáng		
	B_1	B_2	B_3
A_1	5	6	1
	4	8	2
	7	7	3
	6	5	4
A_2	2	6	7
	1	3	10
	3	5	9
	2	4	8

Bài 96. Một nhà kinh doanh cà phê muốn chọn hai hình thức quảng bá cho sản phẩm của mình:

- Quảng cáo trên báo chí.
- Khuyến mãi.

Ông ta cho rằng, trên thị trường đầy cạnh tranh, nếu chỉ quảng cáo thôi thì chưa đủ mà phải đồng thời khuyến mãi để hấp dẫn khách mua hàng. Tuy nhiên, nếu quảng cáo hay khuyến mãi thái quá có thể có tác tác dụng tiêu cực. Do đó, ông thử nghiệm với 3 mức khuyến mãi và 3 mức chi phí quảng cáo (tính bằng % trên doanh thu lúc chưa thử nghiệm). Ông chọn ngẫu nhiên 9 thành phố gần giống nhau để thử nghiệm. Mỗi kết hợp mức giảm giá-mức chi phí quảng cáo

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

được tiến hành ở 1 thành phố. Mức tăng lượng bán 1 tháng thử nghiệm ở mỗi thành phố là như sau (tính bằng 1000kg):

Mức khuyến mãi (sản phẩm)	Chi phí quảng cáo (%)		
	0,6	1	1,4
1	2	7	8
2	3	5	9
3	1	4	7

1. Thiết kế trên thuộc loại gì? Vì sao?
2. Hãy tiến hành một phân tích phương sai để kiểm định sự ảnh hưởng khác nhau của các mức khuyến mãi đến lượng bán với $\alpha = 0,05$;
3. Hãy tiến hành một phân tích phương sai để kiểm định sự ảnh hưởng khác nhau của các mức quảng cáo đến lượng bán với $\alpha = 0,05$;
4. So sánh bội giữa các mức chi phí quảng cáo với $\alpha = 0,05$; (nếu cần)
5. So sánh bội giữa các mức giảm giá với $\alpha = 0,05$; (nếu cần)

BẢNG SỐ NGẪU NHIÊN
(Hàm excel: RANDBETWEEN)

98007	60604	88995	46617	21299	76827	54872	86804	42946	18281
12302	66207	58455	84225	56139	72372	38832	10730	80956	63240
35310	18934	48775	26423	13996	73656	25748	31675	11008	80459
03418	29615	48741	99359	70162	99905	03897	03214	48445	01917
67910	59011	63561	31138	90344	60073	59395	34269	45964	87426
26911	86322	16047	28349	60952	82699	24570	30308	10978	49486
19904	18336	41386	24982	42671	54039	67022	15680	70309	64312
79504	79778	79064	17920	30003	41334	13974	78057	84103	88260
07480	08966	14304	57158	37297	64864	45930	58269	64171	39140
21900	43529	40144	02072	02130	70952	27961	05139	63091	68392
22208	80059	13202	49892	12860	10663	89688	87640	81829	05771
52449	92764	73122	51161	40715	08005	27305	68740	07929	65465
20811	61110	00635	12461	00754	69579	96915	58754	55910	23783
70022	48100	14414	11029	92840	07404	08878	43507	48781	76522
99487	51454	64760	71831	95734	80438	31855	86743	67443	20853
29301	82238	85629	47804	17417	70568	83889	99399	15113	45253
80935	71709	27580	60491	54860	58422	39817	13083	09812	13355
41594	30076	25663	63952	99475	15134	31303	53850	41557	56865
92209	42723	70522	52168	52886	41438	70617	47752	24824	10694
24232	52937	60576	01282	03662	26875	04599	64052	53676	93319
61492	07733	30549	86023	72207	32743	13157	77972	38936	19083
51778	97504	89224	96481	70183	84463	19971	74847	86718	89547
11963	63256	69906	69909	55507	55501	94293	84201	77825	10370
12073	40129	12860	04785	06195	97708	04077	87933	40504	81268
80862	14441	72600	03400	99231	84878	45015	53948	87399	71490
20331	57109	86840	57500	75881	49910	29518	67983	23609	85022
41838	20444	70800	11151	16001	73293	87219	34095	39369	00241
94333	11563	22437	68859	42232	84945	16108	51305	89862	01077
58800	45454	54207	27223	44099	34468	44044	12299	49532	62969
43889	97534	17078	57976	63762	29658	26112	72649	20411	79788
49293	81030	65584	21039	32148	01883	27006	68126	86853	79772
77700	31932	40947	01456	99203	92297	31526	75820	54399	42061
85589	55330	35524	05542	77105	27284	03000	84539	01514	52882
80706	34999	58199	82272	23322	27491	81597	23701	21381	94577
11582	94906	59627	18213	43718	34834	94308	60360	50023	03848
65072	89807	10858	50554	07035	08493	48601	73315	78674	30305
03659	54640	71151	15574	43541	96570	14743	14338	45512	35115
47945	21543	43547	61971	64867	97256	80136	31541	06635	55068
53615	81472	12320	71770	01636	44768	95074	21445	11374	81637

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

BẢNG PHÂN PHỐI NHỊ THỨC

(Hàm excel: BINOM.DIST với Cumulative = 0)

X∈ B(n, p)

n	k	p			
		0,2	0,3	0,4	0,5
7	0	0,210	0,082	0,028	0,0078
	1	0,367	0,247	0,131	0,0547
	2	0,275	0,318	0,261	0,1641
	3	0,115	0,227	0,290	0,2734
	4	0,029	0,097	0,194	0,2734
	5	0,004	0,025	0,077	0,1641
	6	0,000	0,004	0,017	0,0547
	7	0,000	0,000	0,002	0,0078
8	0	0,168	0,058	0,017	0,0039
	1	0,336	0,198	0,090	0,0313
	2	0,294	0,297	0,209	0,1094
	3	0,147	0,254	0,279	0,2188
	4	0,046	0,136	0,232	0,2734
	5	0,009	0,047	0,124	0,2188
	6	0,001	0,010	0,041	0,1094
	7	0,000	0,001	0,008	0,0313
	8	0,000	0,000	0,001	0,0039
9	0	0,134	0,040	0,010	0,0020
	1	0,302	0,156	0,061	0,0176
	2	0,302	0,267	0,161	0,0703
	3	0,176	0,267	0,251	0,1641
	4	0,066	0,172	0,251	0,2461
	5	0,017	0,074	0,167	0,2461
	6	0,003	0,021	0,074	0,1641
	7	0,000	0,004	0,021	0,0703
	8	0,000	0,000	0,004	0,0176
	9	0,000	0,000	0,000	0,0020
10	0	0,107	0,028	0,006	0,0010
	1	0,268	0,121	0,040	0,0098
	2	0,302	0,234	0,121	0,0439
	3	0,201	0,267	0,215	0,1172
	4	0,088	0,200	0,251	0,2051
10	5	0,026	0,103	0,201	0,2461
	6	0,006	0,037	0,112	0,2051
	7	0,001	0,009	0,043	0,1172
	8	0,000	0,001	0,011	0,0439
	9	0,000	0,000	0,002	0,0098
	10	0,000	0,000	0,000	0,0010
11	0	0,086	0,020	0,004	0,0005
	1	0,236	0,093	0,027	0,0054
	2	0,295	0,200	0,089	0,0269
	3	0,222	0,257	0,177	0,0806
	4	0,111	0,220	0,237	0,1611
	5	0,039	0,132	0,221	0,2256
	6	0,010	0,057	0,147	0,2256
	7	0,002	0,017	0,070	0,1611
	8	0,000	0,004	0,023	0,0806
	9	0,000	0,001	0,005	0,0269
	10	0,000	0,000	0,001	0,0054
	11	0,000	0,000	0,000	0,0005
12	0	0,069	0,014	0,002	0,0002
	1	0,206	0,071	0,017	0,0029
	2	0,284	0,168	0,064	0,0161
	3	0,236	0,240	0,142	0,0537
	4	0,133	0,231	0,213	0,1208
	5	0,053	0,159	0,227	0,1934
	6	0,016	0,079	0,177	0,2256
	7	0,003	0,029	0,101	0,1934
	8	0,001	0,008	0,042	0,1208
	9	0,000	0,002	0,013	0,0537
	10	0,000	0,000	0,003	0,0161
	11	0,000	0,000	0,000	0,0029
	12	0,000	0,000	0,000	0,0002

BẢNG PHÂN PHỐI NHỊ THỨC

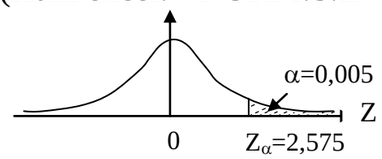
n	k	p			
		0,2	0,3	0,4	0,5
13	0	0,055	0,010	0,001	0,0001
	1	0,179	0,054	0,011	0,0016
	2	0,268	0,139	0,045	0,0095
	3	0,246	0,218	0,111	0,0349
	4	0,154	0,234	0,185	0,0873
	5	0,069	0,180	0,221	0,1571
	6	0,023	0,103	0,197	0,2095
	7	0,006	0,044	0,131	0,2095
	8	0,001	0,014	0,066	0,1571
	9	0,000	0,003	0,024	0,0873
	10	0,000	0,001	0,007	0,0349
	11	0,000	0,000	0,001	0,0095
	12	0,000	0,000	0,000	0,0016
	13	0,000	0,000	0,000	0,0001
14	0	0,044	0,007	0,001	0,0001
	1	0,154	0,041	0,007	0,0009
	2	0,250	0,113	0,032	0,0056
	3	0,250	0,194	0,085	0,0222
	4	0,172	0,229	0,155	0,0611
	5	0,086	0,196	0,207	0,1222
	6	0,032	0,126	0,207	0,1833
	7	0,009	0,062	0,157	0,2095
	8	0,002	0,023	0,092	0,1833
	9	0,000	0,007	0,041	0,1222
	10	0,000	0,001	0,014	0,0611
	11	0,000	0,000	0,003	0,0222
	12	0,000	0,000	0,001	0,0056
	13	0,000	0,000	0,000	0,0009
	14	0,000	0,000	0,000	0,0001
15	0	0,035	0,005	0,001	0,0000
	1	0,132	0,031	0,005	0,0005
	2	0,231	0,092	0,022	0,0032
	3	0,250	0,170	0,063	0,0139
	4	0,188	0,219	0,127	0,0417
	5	0,103	0,206	0,186	0,0916
15	6	0,043	0,147	0,207	0,1527
	7	0,014	0,081	0,177	0,1964
	8	0,004	0,035	0,118	0,1964
	9	0,001	0,012	0,061	0,1527
	10	0,000	0,003	0,025	0,0916
	11	0,000	0,001	0,007	0,0417
	12	0,000	0,000	0,002	0,0139
	13	0,000	0,000	0,000	0,0032
	14	0,000	0,000	0,000	0,0005
	15	0,000	0,000	0,000	0,0000
16	0	0,028	0,003	0,000	0,0000
	1	0,113	0,023	0,003	0,0002
	2	0,211	0,073	0,015	0,0018
	3	0,246	0,147	0,047	0,0085
	4	0,200	0,204	0,101	0,0278
	5	0,120	0,210	0,162	0,0667
	6	0,055	0,165	0,198	0,1222
	7	0,020	0,101	0,189	0,1746
	8	0,006	0,049	0,142	0,1964
	9	0,001	0,019	0,084	0,1746
	10	0,000	0,006	0,039	0,1222
	11	0,000	0,001	0,014	0,0667
	12	0,000	0,000	0,004	0,0278
	13	0,000	0,000	0,001	0,0085
	14	0,000	0,000	0,000	0,0018
	15	0,000	0,000	0,000	0,0002
	16	0,000	0,000	0,000	0,0000
17	0	0,023	0,002	0,000	0,0000
	1	0,096	0,017	0,002	0,0001
	2	0,191	0,058	0,010	0,0010
	3	0,239	0,125	0,034	0,0052
	4	0,209	0,187	0,080	0,0182
	5	0,136	0,208	0,138	0,0472
	6	0,068	0,178	0,184	0,0944
	7	0,027	0,120	0,193	0,1484

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

BẢNG PHÂN VỊ CHUẨN

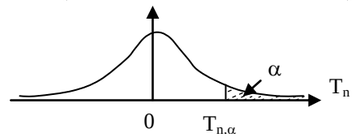
(Hàm excel: - NORM.S.INV)



α	Z_α	α	Z_α	α	Z_α	α	Z_α
0,0000	4,000	***0,01***	2,326	***0,025***	1,960	***0,05***	1,645
0,0002	3,500	0,0106	2,305	0,0253	1,955	0,0548	1,600
0,0013	3,000	0,0112	2,285	0,0256	1,950	0,0606	1,550
0,005	2,575	0,0118	2,270	0,0262	1,940	0,0668	1,500
0,0062	2,500	0,0124	2,245	0,0268	1,930	0,0735	1,450
0,0064	2,490	0,0125	2,240	0,0274	1,920	0,0808	1,400
0,0066	2,480	0,0129	2,230	0,0281	1,910	0,0885	1,350
0,0068	2,470	0,0132	2,220	0,0287	1,900	0,0968	1,300
0,0069	2,460	0,0136	2,210	0,0294	1,890	0,1056	1,250
0,0071	2,450	0,0139	2,200	0,0301	1,880	0,1151	1,200
0,0072	2,445	0,0143	2,190	0,0307	1,870	0,1251	1,150
0,0073	2,440	0,0146	2,180	0,0314	1,860	0,1357	1,100
0,0074	2,435	0,0150	2,170	0,0322	1,850	0,1469	1,050
0,0075	2,430	0,0154	2,160	0,0329	1,840	0,1587	1,000
0,0077	2,425	0,0158	2,150	0,0336	1,830	0,1841	0,900
0,0078	2,420	0,0162	2,140	0,0344	1,820	0,2119	0,800
0,0079	2,415	0,0166	2,130	0,0351	1,810	0,2420	0,700
0,0080	2,410	0,0170	2,120	0,0359	1,800	0,2743	0,600
0,0081	2,405	0,0174	2,110	0,0367	1,790	0,3085	0,500
0,0082	2,400	0,0179	2,100	0,0375	1,780	0,3446	0,400
0,0083	2,395	0,0183	2,090	0,0384	1,770	0,3821	0,300
0,0084	2,390	0,0188	2,080	0,0392	1,760	0,4207	0,200
0,0085	2,385	0,0192	2,070	0,0401	1,750	0,4602	0,100
0,0087	2,380	0,0197	2,060	0,0409	1,740	0,4641	0,090
0,0088	2,375	0,0202	2,050	0,0418	1,730	0,4681	0,080
0,0089	2,370	0,0207	2,040	0,0427	1,720	0,4721	0,070
0,0090	2,365	0,0212	2,030	0,0436	1,710	0,4761	0,060
0,0091	2,360	0,0217	2,020	0,0446	1,700	0,4801	0,050
0,0093	2,355	0,0222	2,010	0,0455	1,690	0,4840	0,040
0,0094	2,350	0,0228	2,000	0,0465	1,680	0,4880	0,030
0,0095	2,345	0,0233	1,990	0,0475	1,670	0,4920	0,020
0,0096	2,340	0,0239	1,980	0,0485	1,660	0,4960	0,010
0,0099	2,330	0,0244	1,970	0,0495	1,650	0,5000	0,000

BẢNG PHÂN VỊ STUDENT

(Hàm excel: - T.INV)

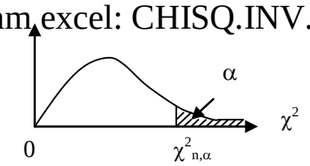


n	***α***					
	0,1	***0,05***	***0,025***	***0,01***	***0,005***	***0,001***
1	3,078	6,314	12,706	31,821	63,657	318,309
2	1,886	2,920	4,303	6,965	9,925	22,327
3	1,638	2,353	3,182	4,541	5,841	10,215
4	1,533	2,132	2,776	3,747	4,604	7,173
5	1,476	2,015	2,571	3,365	4,032	5,893
6	1,440	1,943	2,447	3,143	3,707	5,208
7	1,415	1,895	2,365	2,998	3,499	4,785
8	1,397	1,860	2,306	2,896	3,355	4,501
9	1,383	1,833	2,262	2,821	3,250	4,297
10	1,372	1,812	2,228	2,764	3,169	4,144
11	1,363	1,796	2,201	2,718	3,106	4,025
12	1,356	1,782	2,179	2,681	3,055	3,930
13	1,350	1,771	2,160	2,650	3,012	3,852
14	1,345	1,761	2,145	2,624	2,977	3,787
15	1,341	1,753	2,131	2,602	2,947	3,733
16	1,337	1,746	2,120	2,583	2,921	3,686
17	1,333	1,740	2,110	2,567	2,898	3,646
18	1,330	1,734	2,101	2,552	2,878	3,610
19	1,328	1,729	2,093	2,539	2,861	3,579
20	1,325	1,725	2,086	2,528	2,845	3,552
21	1,323	1,721	2,080	2,518	2,831	3,527
22	1,321	1,717	2,074	2,508	2,819	3,505
23	1,319	1,714	2,069	2,500	2,807	3,485
24	1,318	1,711	2,064	2,492	2,797	3,467
25	1,316	1,708	2,060	2,485	2,787	3,450
26	1,315	1,706	2,056	2,479	2,779	3,435
27	1,314	1,703	2,052	2,473	2,771	3,421
28	1,313	1,701	2,048	2,467	2,763	3,408
29	1,311	1,699	2,045	2,462	2,756	3,396
30	1,310	1,697	2,042	2,457	2,750	3,385
40	1,303	1,684	2,021	2,423	2,704	3,307
60	1,296	1,671	2,000	2,390	2,660	3,232
120	1,289	1,658	1,980	2,358	2,617	3,160
inf	1,282	1,645	1,960	2,326	2,576	3,090

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

BẢNG PHÂN VỊ KHI BÌNH PHƯƠNG

(Hàm excel: CHISQ.INV.RT)



n	*α*							
	0,995	***0,99***	***0,975***	***0,95***	***0,05***	***0,025***	***0,01***	***0,005***
1	0,000	0,000	0,001	0,004	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	36,415	39,364	42,980	45,559
25	10,520	11,524	13,120	14,611	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	42,557	45,722	49,588	52,336
30	13,787	14,953	16,791	18,493	43,773	46,979	50,892	53,672
40	20,707	22,164	24,433	26,509	55,758	59,342	63,691	66,766
50	27,991	29,707	32,357	34,764	67,505	71,420	76,154	79,490
60	35,534	37,485	40,482	43,188	79,082	83,298	88,379	91,952
120	83,852	86,923	91,573	95,705	146,57	152,21	158,95	163,65

BẢNG PHÂN VỊ FISHER

(Hàm excel: F.INV.RT)

α = 0,01

n_1	n_2											
	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***
1	4052	98,50	34,12	21,20	16,26	13,75	12,3	11,3	10,6	10,0	9,65	9,33
2	4999	99,00	30,82	18,00	13,27	10,92	9,55	8,65	8,02	7,56	7,21	6,93
3	5403	99,17	29,46	16,69	12,06	9,78	8,45	7,59	6,99	6,55	6,22	5,95
4	5625	99,25	28,71	15,98	11,39	9,15	7,85	7,01	6,42	5,99	5,67	5,41
5	5764	99,30	28,24	15,52	10,97	8,75	7,46	6,63	6,06	5,64	5,32	5,06
6	5859	99,33	27,91	15,21	10,67	8,47	7,19	6,37	5,80	5,39	5,07	4,82
7	5928	99,36	27,67	14,98	10,46	8,26	6,99	6,18	5,61	5,20	4,89	4,64
8	5981	99,37	27,49	14,80	10,29	8,10	6,84	6,03	5,47	5,06	4,74	4,50
9	6022	99,39	27,35	14,66	10,16	7,98	6,72	5,91	5,35	4,94	4,63	4,39
10	6056	99,40	27,23	14,55	10,05	7,87	6,62	5,81	5,26	4,85	4,54	4,30
11	6083	99,41	27,13	14,45	9,96	7,79	6,54	5,73	5,18	4,77	4,46	4,22
12	6106	99,42	27,05	14,37	9,89	7,72	6,47	5,67	5,11	4,71	4,40	4,16
13	6126	99,42	26,98	14,31	9,82	7,66	6,41	5,61	5,05	4,65	4,34	4,10
14	6143	99,43	26,92	14,25	9,77	7,60	6,36	5,56	5,01	4,60	4,29	4,05
15	6157	99,43	26,87	14,20	9,72	7,56	6,31	5,52	4,96	4,56	4,25	4,01
16	6170	99,44	26,83	14,15	9,68	7,52	6,28	5,48	4,92	4,52	4,21	3,97
17	6181	99,44	26,79	14,11	9,64	7,48	6,24	5,44	4,89	4,49	4,18	3,94
18	6192	99,44	26,75	14,08	9,61	7,45	6,21	5,41	4,86	4,46	4,15	3,91
19	6201	99,45	26,72	14,05	9,58	7,42	6,18	5,38	4,83	4,43	4,12	3,88
20	6209	99,45	26,69	14,02	9,55	7,40	6,16	5,36	4,81	4,41	4,10	3,86
21	6216	99,45	26,66	13,99	9,53	7,37	6,13	5,34	4,79	4,38	4,08	3,84
22	6223	99,45	26,64	13,97	9,51	7,35	6,11	5,32	4,77	4,36	4,06	3,82
23	6229	99,46	26,62	13,95	9,49	7,33	6,09	5,30	4,75	4,34	4,04	3,80
24	6235	99,46	26,60	13,93	9,47	7,31	6,07	5,28	4,73	4,33	4,02	3,78
25	6240	99,46	26,58	13,91	9,45	7,30	6,06	5,26	4,71	4,31	4,01	3,76
26	6245	99,46	26,56	13,89	9,43	7,28	6,04	5,25	4,70	4,30	3,99	3,75
27	6249	99,46	26,55	13,88	9,42	7,27	6,03	5,23	4,68	4,28	3,98	3,74
28	6253	99,46	26,53	13,86	9,40	7,25	6,02	5,22	4,67	4,27	3,96	3,72
29	6257	99,46	26,52	13,85	9,39	7,24	6,00	5,21	4,66	4,26	3,95	3,71
30	6261	99,47	26,50	13,84	9,38	7,23	5,99	5,20	4,65	4,25	3,94	3,70
40	6287	99,47	26,41	13,75	9,29	7,14	5,91	5,12	4,57	4,17	3,86	3,62
60	6313	99,48	26,32	13,65	9,20	7,06	5,82	5,03	4,48	4,08	3,78	3,54
120	6339	99,49	26,22	13,56	9,11	6,97	5,74	4,95	4,40	4,00	3,69	3,45
inf	6366	99,50	26,13	13,46	9,02	6,88	5,65	4,86	4,31	3,91	3,60	3,36

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

BẢNG PHÂN VỊ FISHER

α = 0,01

n_1	n_2											
	13	***14***	***16***	***20***	***24***	***27***	***30***	***40***	***60***	***70***	***120***	***inf***
1	9,07	8,86	8,53	8,10	7,82	7,68	7,56	7,31	7,08	7,01	6,85	6,63
2	6,70	6,51	6,23	5,85	5,61	5,49	5,39	5,18	4,98	4,92	4,79	4,61
3	5,74	5,56	5,29	4,94	4,72	4,60	4,51	4,31	4,13	4,07	3,95	3,78
4	5,21	5,04	4,77	4,43	4,22	4,11	4,02	3,83	3,65	3,60	3,48	3,32
5	4,86	4,69	4,44	4,10	3,90	3,78	3,70	3,51	3,34	3,29	3,17	3,02
6	4,62	4,46	4,20	3,87	3,67	3,56	3,47	3,29	3,12	3,07	2,96	2,80
7	4,44	4,28	4,03	3,70	3,50	3,39	3,30	3,12	2,95	2,91	2,79	2,64
8	4,30	4,14	3,89	3,56	3,36	3,26	3,17	2,99	2,82	2,78	2,66	2,51
9	4,19	4,03	3,78	3,46	3,26	3,15	3,07	2,89	2,72	2,67	2,56	2,41
10	4,10	3,94	3,69	3,37	3,17	3,06	2,98	2,80	2,63	2,59	2,47	2,32
11	4,02	3,86	3,62	3,29	3,09	2,99	2,91	2,73	2,56	2,51	2,40	2,25
12	3,96	3,80	3,55	3,23	3,03	2,93	2,84	2,66	2,50	2,45	2,34	2,18
13	3,91	3,75	3,50	3,18	2,98	2,87	2,79	2,61	2,44	2,40	2,28	2,13
14	3,86	3,70	3,45	3,13	2,93	2,82	2,74	2,56	2,39	2,35	2,23	2,08
15	3,82	3,66	3,41	3,09	2,89	2,78	2,70	2,52	2,35	2,31	2,19	2,04
16	3,78	3,62	3,37	3,05	2,85	2,75	2,66	2,48	2,31	2,27	2,15	2,00
17	3,75	3,59	3,34	3,02	2,82	2,71	2,63	2,45	2,28	2,23	2,12	1,97
18	3,72	3,56	3,31	2,99	2,79	2,68	2,60	2,42	2,25	2,20	2,09	1,93
19	3,69	3,53	3,28	2,96	2,76	2,66	2,57	2,39	2,22	2,18	2,06	1,90
20	3,66	3,51	3,26	2,94	2,74	2,63	2,55	2,37	2,20	2,15	2,03	1,88
21	3,64	3,48	3,24	2,92	2,72	2,61	2,53	2,35	2,17	2,13	2,01	1,85
22	3,62	3,46	3,22	2,90	2,70	2,59	2,51	2,33	2,15	2,11	1,99	1,83
23	3,60	3,44	3,20	2,88	2,68	2,57	2,49	2,31	2,13	2,09	1,97	1,81
24	3,59	3,43	3,18	2,86	2,66	2,55	2,47	2,29	2,12	2,07	1,95	1,79
25	3,57	3,41	3,16	2,84	2,64	2,54	2,45	2,27	2,10	2,05	1,93	1,77
26	3,56	3,40	3,15	2,83	2,63	2,52	2,44	2,26	2,08	2,03	1,92	1,76
27	3,54	3,38	3,14	2,81	2,61	2,51	2,42	2,24	2,07	2,02	1,90	1,74
28	3,53	3,37	3,12	2,80	2,60	2,49	2,41	2,23	2,05	2,01	1,89	1,72
29	3,52	3,36	3,11	2,79	2,59	2,48	2,40	2,22	2,04	1,99	1,87	1,71
30	3,51	3,35	3,10	2,78	2,58	2,47	2,39	2,20	2,03	1,98	1,86	1,70
40	3,43	3,27	3,02	2,69	2,49	2,38	2,30	2,11	1,94	1,89	1,76	1,59
50	3,38	3,22	2,97	2,64	2,44	2,33	2,25	2,06	1,88	1,83	1,70	1,52
60	3,34	3,18	2,93	2,61	2,40	2,29	2,21	2,02	1,84	1,78	1,66	1,47
120	3,25	3,09	2,84	2,52	2,31	2,20	2,11	1,92	1,73	1,67	1,53	1,32
inf	3,17	3,00	2,75	2,42	2,21	2,10	2,01	1,80	1,60	1,54	1,38	1,00

BẢNG PHÂN VỊ FISHER

α = 0,025

n_1	n_2											
	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***
1	648	38,51	17,44	12,22	10,01	8,81	8,07	7,57	7,21	6,94	6,72	6,55
2	799	39,00	16,04	10,65	8,43	7,26	6,54	6,06	5,71	5,46	5,26	5,10
3	864	39,17	15,44	9,98	7,76	6,60	5,89	5,42	5,08	4,83	4,63	4,47
4	900	39,25	15,10	9,60	7,39	6,23	5,52	5,05	4,72	4,47	4,28	4,12
5	922	39,30	14,88	9,36	7,15	5,99	5,29	4,82	4,48	4,24	4,04	3,89
6	937	39,33	14,73	9,20	6,98	5,82	5,12	4,65	4,32	4,07	3,88	3,73
7	948	39,36	14,62	9,07	6,85	5,70	4,99	4,53	4,20	3,95	3,76	3,61
8	957	39,37	14,54	8,98	6,76	5,60	4,90	4,43	4,10	3,85	3,66	3,51
9	963	39,39	14,47	8,90	6,68	5,52	4,82	4,36	4,03	3,78	3,59	3,44
10	969	39,40	14,42	8,84	6,62	5,46	4,76	4,30	3,96	3,72	3,53	3,37
11	973	39,41	14,37	8,79	6,57	5,41	4,71	4,24	3,91	3,66	3,47	3,32
12	977	39,41	14,34	8,75	6,52	5,37	4,67	4,20	3,87	3,62	3,43	3,28
13	980	39,42	14,30	8,71	6,49	5,33	4,63	4,16	3,83	3,58	3,39	3,24
14	983	39,43	14,28	8,68	6,46	5,30	4,60	4,13	3,80	3,55	3,36	3,21
15	985	39,43	14,25	8,66	6,43	5,27	4,57	4,10	3,77	3,52	3,33	3,18
16	987	39,44	14,23	8,63	6,40	5,24	4,54	4,08	3,74	3,50	3,30	3,15
17	989	39,44	14,21	8,61	6,38	5,22	4,52	4,05	3,72	3,47	3,28	3,13
18	990	39,44	14,20	8,59	6,36	5,20	4,50	4,03	3,70	3,45	3,26	3,11
19	992	39,45	14,18	8,58	6,34	5,18	4,48	4,02	3,68	3,44	3,24	3,09
20	993	39,45	14,17	8,56	6,33	5,17	4,47	4,00	3,67	3,42	3,23	3,07
21	994	39,45	14,16	8,55	6,31	5,15	4,45	3,98	3,65	3,40	3,21	3,06
22	995	39,45	14,14	8,53	6,30	5,14	4,44	3,97	3,64	3,39	3,20	3,04
23	996	39,45	14,13	8,52	6,29	5,13	4,43	3,96	3,63	3,38	3,18	3,03
24	997	39,46	14,12	8,51	6,28	5,12	4,41	3,95	3,61	3,37	3,17	3,02
25	998	39,46	14,12	8,50	6,27	5,11	4,40	3,94	3,60	3,35	3,16	3,01
26	999	39,46	14,11	8,49	6,26	5,10	4,39	3,93	3,59	3,34	3,15	3,00
27	1000	39,46	14,10	8,48	6,25	5,09	4,39	3,92	3,58	3,34	3,14	2,99
28	1000	39,46	14,09	8,48	6,24	5,08	4,38	3,91	3,58	3,33	3,13	2,98
29	1001	39,46	14,09	8,47	6,23	5,07	4,37	3,90	3,57	3,32	3,13	2,97
30	1001	39,46	14,08	8,46	6,23	5,07	4,36	3,89	3,56	3,31	3,12	2,96
40	1006	39,47	14,04	8,41	6,18	5,01	4,31	3,84	3,51	3,26	3,06	2,91
60	1010	39,48	13,99	8,36	6,12	4,96	4,25	3,78	3,45	3,20	3,00	2,85
120	1014	39,49	13,95	8,31	6,07	4,90	4,20	3,73	3,39	3,14	2,94	2,79
inf	1018	39,50	13,90	8,26	6,02	4,85	4,14	3,67	3,33	3,08	2,88	2,72

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

BẢNG PHÂN VỊ FISHER

α = 0,025

n_1	n_2											
	13	***14***	***16***	***20***	***24***	***27***	***30***	***40***	***60***	***70***	***120***	***inf***
1	6,41	6,30	6,12	5,87	5,72	5,63	5,57	5,42	5,29	5,25	5,15	5,02
2	4,97	4,86	4,69	4,46	4,32	4,24	4,18	4,05	3,93	3,89	3,80	3,69
3	4,35	4,24	4,08	3,86	3,72	3,65	3,59	3,46	3,34	3,31	3,23	3,12
4	4,00	3,89	3,73	3,51	3,38	3,31	3,25	3,13	3,01	2,97	2,89	2,79
5	3,77	3,66	3,50	3,29	3,15	3,08	3,03	2,90	2,79	2,75	2,67	2,57
6	3,60	3,50	3,34	3,13	2,99	2,92	2,87	2,74	2,63	2,59	2,52	2,41
7	3,48	3,38	3,22	3,01	2,87	2,80	2,75	2,62	2,51	2,47	2,39	2,29
8	3,39	3,29	3,12	2,91	2,78	2,71	2,65	2,53	2,41	2,38	2,30	2,19
9	3,31	3,21	3,05	2,84	2,70	2,63	2,57	2,45	2,33	2,30	2,22	2,11
10	3,25	3,15	2,99	2,77	2,64	2,57	2,51	2,39	2,27	2,24	2,16	2,05
11	3,20	3,09	2,93	2,72	2,59	2,51	2,46	2,33	2,22	2,18	2,10	1,99
12	3,15	3,05	2,89	2,68	2,54	2,47	2,41	2,29	2,17	2,14	2,05	1,94
13	3,12	3,01	2,85	2,64	2,50	2,43	2,37	2,25	2,13	2,10	2,01	1,90
14	3,08	2,98	2,82	2,60	2,47	2,39	2,34	2,21	2,09	2,06	1,98	1,87
15	3,05	2,95	2,79	2,57	2,44	2,36	2,31	2,18	2,06	2,03	1,94	1,83
16	3,03	2,92	2,76	2,55	2,41	2,34	2,28	2,15	2,03	2,00	1,92	1,80
17	3,00	2,90	2,74	2,52	2,39	2,31	2,26	2,13	2,01	1,97	1,89	1,78
18	2,98	2,88	2,72	2,50	2,36	2,29	2,23	2,11	1,98	1,95	1,87	1,75
19	2,96	2,86	2,70	2,48	2,35	2,27	2,21	2,09	1,96	1,93	1,84	1,73
20	2,95	2,84	2,68	2,46	2,33	2,25	2,20	2,07	1,94	1,91	1,82	1,71
21	2,93	2,83	2,67	2,45	2,31	2,24	2,18	2,05	1,93	1,89	1,81	1,69
22	2,92	2,81	2,65	2,43	2,30	2,22	2,16	2,03	1,91	1,88	1,79	1,67
23	2,91	2,80	2,64	2,42	2,28	2,21	2,15	2,02	1,90	1,86	1,77	1,66
24	2,89	2,79	2,63	2,41	2,27	2,19	2,14	2,01	1,88	1,85	1,76	1,64
25	2,88	2,78	2,61	2,40	2,26	2,18	2,12	1,99	1,87	1,83	1,75	1,63
26	2,87	2,77	2,60	2,39	2,25	2,17	2,11	1,98	1,86	1,82	1,73	1,61
27	2,86	2,76	2,59	2,38	2,24	2,16	2,10	1,97	1,85	1,81	1,72	1,60
28	2,85	2,75	2,58	2,37	2,23	2,15	2,09	1,96	1,83	1,80	1,71	1,59
29	2,85	2,74	2,58	2,36	2,22	2,14	2,08	1,95	1,82	1,79	1,70	1,58
30	2,84	2,73	2,57	2,35	2,21	2,13	2,07	1,94	1,82	1,78	1,69	1,57
40	2,78	2,67	2,51	2,29	2,15	2,07	2,01	1,88	1,74	1,71	1,61	1,48
50	2,74	2,64	2,47	2,25	2,11	2,03	1,97	1,83	1,70	1,66	1,56	1,43
60	2,72	2,61	2,45	2,22	2,08	2,00	1,94	1,80	1,67	1,63	1,53	1,39
120	2,66	2,55	2,38	2,16	2,01	1,93	1,87	1,72	1,58	1,54	1,43	1,27
inf	2,60	2,49	2,32	2,09	1,94	1,85	1,79	1,64	1,48	1,44	1,31	1,00

BẢNG PHÂN VỊ FISHER

α = 0,05

n_1	n_2											
	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***
1	161,4	18,51	10,13	7,71	6,61	5,99	5,59	5,32	5,12	4,96	4,84	4,75
2	199,5	19,00	9,55	6,94	5,79	5,14	4,74	4,46	4,26	4,10	3,98	3,89
3	215,7	19,16	9,28	6,59	5,41	4,76	4,35	4,07	3,86	3,71	3,59	3,49
4	224,6	19,25	9,12	6,39	5,19	4,53	4,12	3,84	3,63	3,48	3,36	3,26
5	230,2	19,30	9,01	6,26	5,05	4,39	3,97	3,69	3,48	3,33	3,20	3,11
6	234,0	19,33	8,94	6,16	4,95	4,28	3,87	3,58	3,37	3,22	3,09	3,00
7	236,8	19,35	8,89	6,09	4,88	4,21	3,79	3,50	3,29	3,14	3,01	2,91
8	238,9	19,37	8,85	6,04	4,82	4,15	3,73	3,44	3,23	3,07	2,95	2,85
9	240,5	19,38	8,81	6,00	4,77	4,10	3,68	3,39	3,18	3,02	2,90	2,80
10	241,9	19,40	8,79	5,96	4,74	4,06	3,64	3,35	3,14	2,98	2,85	2,75
11	243,0	19,40	8,76	5,94	4,70	4,03	3,60	3,31	3,10	2,94	2,82	2,72
12	243,9	19,41	8,74	5,91	4,68	4,00	3,57	3,28	3,07	2,91	2,79	2,69
13	244,7	19,42	8,73	5,89	4,66	3,98	3,55	3,26	3,05	2,89	2,76	2,66
14	245,4	19,42	8,71	5,87	4,64	3,96	3,53	3,24	3,03	2,86	2,74	2,64
15	245,9	19,43	8,70	5,86	4,62	3,94	3,51	3,22	3,01	2,85	2,72	2,62
16	246,5	19,43	8,69	5,84	4,60	3,92	3,49	3,20	2,99	2,83	2,70	2,60
17	246,9	19,44	8,68	5,83	4,59	3,91	3,48	3,19	2,97	2,81	2,69	2,58
18	247,3	19,44	8,67	5,82	4,58	3,90	3,47	3,17	2,96	2,80	2,67	2,57
19	247,7	19,44	8,67	5,81	4,57	3,88	3,46	3,16	2,95	2,79	2,66	2,56
20	248,0	19,45	8,66	5,80	4,56	3,87	3,44	3,15	2,94	2,77	2,65	2,54
21	248,3	19,45	8,65	5,79	4,55	3,86	3,43	3,14	2,93	2,76	2,64	2,53
22	248,6	19,45	8,65	5,79	4,54	3,86	3,43	3,13	2,92	2,75	2,63	2,52
23	248,8	19,45	8,64	5,78	4,53	3,85	3,42	3,12	2,91	2,75	2,62	2,51
24	249,1	19,45	8,64	5,77	4,53	3,84	3,41	3,12	2,90	2,74	2,61	2,51
25	249,3	19,46	8,63	5,77	4,52	3,83	3,40	3,11	2,89	2,73	2,60	2,50
26	249,5	19,46	8,63	5,76	4,52	3,83	3,40	3,10	2,89	2,72	2,59	2,49
27	249,6	19,46	8,63	5,76	4,51	3,82	3,39	3,10	2,88	2,72	2,59	2,48
28	249,8	19,46	8,62	5,75	4,50	3,82	3,39	3,09	2,87	2,71	2,58	2,48
29	250,0	19,46	8,62	5,75	4,50	3,81	3,38	3,08	2,87	2,70	2,58	2,47
30	250,1	19,46	8,62	5,75	4,50	3,81	3,38	3,08	2,86	2,70	2,57	2,47
40	251,1	19,47	8,59	5,72	4,46	3,77	3,34	3,04	2,83	2,66	2,53	2,43
60	252,2	19,48	8,57	5,69	4,43	3,74	3,30	3,01	2,79	2,62	2,49	2,38
120	253,3	19,49	8,55	5,66	4,40	3,70	3,27	2,97	2,75	2,58	2,45	2,34
inf	254,3	19,50	8,53	5,63	4,37	3,67	3,23	2,93	2,71	2,54	2,40	2,30

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

BẢNG PHÂN VỊ FISHER

α = 0,05

n_1	n_2											
	13	***14***	***16***	***20***	***24***	***27***	***30***	***40***	***60***	***70***	***120***	***inf***
1	4,67	4,60	4,49	4,35	4,26	4,21	4,17	4,08	4,00	3,98	3,92	3,84
2	3,81	3,74	3,63	3,49	3,40	3,35	3,32	3,23	3,15	3,13	3,07	3,00
3	3,41	3,34	3,24	3,10	3,01	2,96	2,92	2,84	2,76	2,74	2,68	2,60
4	3,18	3,11	3,01	2,87	2,78	2,73	2,69	2,61	2,53	2,50	2,45	2,37
5	3,03	2,96	2,85	2,71	2,62	2,57	2,53	2,45	2,37	2,35	2,29	2,21
6	2,92	2,85	2,74	2,60	2,51	2,46	2,42	2,34	2,25	2,23	2,18	2,10
7	2,83	2,76	2,66	2,51	2,42	2,37	2,33	2,25	2,17	2,14	2,09	2,01
8	2,77	2,70	2,59	2,45	2,36	2,31	2,27	2,18	2,10	2,07	2,02	1,94
9	2,71	2,65	2,54	2,39	2,30	2,25	2,21	2,12	2,04	2,02	1,96	1,88
10	2,67	2,60	2,49	2,35	2,25	2,20	2,16	2,08	1,99	1,97	1,91	1,83
11	2,63	2,57	2,46	2,31	2,22	2,17	2,13	2,04	1,95	1,93	1,87	1,79
12	2,60	2,53	2,42	2,28	2,18	2,13	2,09	2,00	1,92	1,89	1,83	1,75
13	2,58	2,51	2,40	2,25	2,15	2,10	2,06	1,97	1,89	1,86	1,80	1,72
14	2,55	2,48	2,37	2,22	2,13	2,08	2,04	1,95	1,86	1,84	1,78	1,69
15	2,53	2,46	2,35	2,20	2,11	2,06	2,01	1,92	1,84	1,81	1,75	1,67
16	2,51	2,44	2,33	2,18	2,09	2,04	1,99	1,90	1,82	1,79	1,73	1,64
17	2,50	2,43	2,32	2,17	2,07	2,02	1,98	1,89	1,80	1,77	1,71	1,62
18	2,48	2,41	2,30	2,15	2,05	2,00	1,96	1,87	1,78	1,75	1,69	1,60
19	2,47	2,40	2,29	2,14	2,04	1,99	1,95	1,85	1,76	1,74	1,67	1,59
20	2,46	2,39	2,28	2,12	2,03	1,97	1,93	1,84	1,75	1,72	1,66	1,57
21	2,45	2,38	2,26	2,11	2,01	1,96	1,92	1,83	1,73	1,71	1,64	1,56
22	2,44	2,37	2,25	2,10	2,00	1,95	1,91	1,81	1,72	1,70	1,63	1,54
23	2,43	2,36	2,24	2,09	1,99	1,94	1,90	1,80	1,71	1,68	1,62	1,53
24	2,42	2,35	2,24	2,08	1,98	1,93	1,89	1,79	1,70	1,67	1,61	1,52
25	2,41	2,34	2,23	2,07	1,97	1,92	1,88	1,78	1,69	1,66	1,60	1,51
26	2,41	2,33	2,22	2,07	1,97	1,91	1,87	1,77	1,68	1,65	1,59	1,50
27	2,40	2,33	2,21	2,06	1,96	1,90	1,86	1,77	1,67	1,65	1,58	1,49
28	2,39	2,32	2,21	2,05	1,95	1,90	1,85	1,76	1,66	1,64	1,57	1,48
29	2,39	2,31	2,20	2,05	1,95	1,89	1,85	1,75	1,66	1,63	1,56	1,47
30	2,38	2,31	2,19	2,04	1,94	1,88	1,84	1,74	1,65	1,62	1,55	1,46
40	2,34	2,27	2,15	1,99	1,89	1,84	1,79	1,69	1,59	1,57	1,50	1,39
50	2,31	2,24	2,12	1,97	1,86	1,81	1,76	1,66	1,56	1,53	1,46	1,35
60	2,30	2,22	2,11	1,95	1,84	1,79	1,74	1,64	1,53	1,50	1,43	1,32
120	2,25	2,18	2,06	1,90	1,79	1,73	1,68	1,58	1,47	1,44	1,35	1,22
inf	2,21	2,13	2,01	1,84	1,73	1,67	1,62	1,51	1,39	1,35	1,25	1,00

BẢNG PHÂN VỊ WILCOXON

n	***α***			
	0,005	***0,01***	***0,025***	***0,05***
5	0	0	0	1
6	0	0	1	3
7	0	1	3	4
8	1	2	4	6
9	2	4	6	9
10	4	6	9	11
11	6	8	11	14
12	8	10	14	18
13	10	13	18	22
14	13	16	22	26
15	16	20	26	31
16	20	24	30	36
17	24	28	35	42
18	28	33	41	48
19	33	38	47	54
20	38	44	53	61
21	43	50	59	68
22	49	56	66	76
23	55	63	74	84
24	62	70	82	92
25	69	77	90	101
26	76	85	99	111
27	84	93	108	120
28	92	102	117	131
29	101	111	127	141
30	110	121	138	152

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

BẢNG PHÂN VỊ MAN-WHITNEY $\alpha = 0{,}025$

n_2	n_1															
	5	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***
5	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

BẢNG PHÂN VỊ MAN-WHITNEY $\alpha = 0{,}05$

n_2	n_1															
	5	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***
5	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
6	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32
7	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
8	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
9	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
10	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62
11	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69
12	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77
13	15	19	24	28	33	37	42	47	51	56	61	65	70	75	80	84
14	16	21	26	31	36	41	46	51	56	61	66	71	77	82	87	92
15	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100
16	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107
17	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115
18	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123
19	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
20	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138

BẢNG PHÂN VỊ SPEARMAN

n	***α***			
	0,005	***0,01***	***0,025***	***0,05***
5	-	-	-	0,900
6	-	0,943	0,886	0,829
7	-	0,893	0,786	0,714
8	0,881	0,833	0,738	0,643
9	0,833	0,783	0,683	0,600
10	0,794	0,745	0,648	0,564
11	0,818	0,736	0,623	0,523
12	0,780	0,703	0,591	0,496
13	0,745	0,673	0,566	0,475
14	0,716	0,646	0,545	0,457
15	0,689	0,623	0,525	0,441
16	0,666	0,601	0,507	0,425
17	0,645	0,582	0,490	0,412
18	0,625	0,564	0,476	0,399
19	0,608	0,549	0,462	0,388
20	0,591	0,534	0,450	0,377
21	0,576	0,521	0,438	0,368
22	0,562	0,508	0,428	0,359
23	0,549	0,496	0,418	0,351
24	0,537	0,485	0,409	0,343
25	0,526	0,475	0,400	0,336
26	0,515	0,465	0,392	0,329
27	0,505	0,456	0,385	0,323
28	0,496	0,448	0,377	0,317
29	0,487	0,440	0,370	0,311
30	0,478	0,432	0,364	0,305

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

PHÂN PHỐI KHOẢNG STUDENTIZED

$\alpha = 0{,}01$

r_2	r_1									
	2	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***
1	90	135	164	186	202	216	227	237	246	253
2	14,0	19,0	22,3	24,7	26,6	28,2	29,5	30,7	31,7	32,6
3	8,26	10,60	12,20	13,30	14,20	15,00	15,60	16,20	16,70	17,10
4	6,51	8,12	9,17	9,96	10,60	11,10	11,50	11,90	12,30	12,60
5	5,70	6,97	7,80	8,42	8,91	9,32	9,67	9,97	10,20	10,50
6	5,24	6,33	7,03	7,56	7,97	8,32	8,61	8,87	9,10	9,30
7	4,95	5,92	6,54	7,01	7,37	7,68	7,94	8,17	8,37	8,55
8	4,74	5,63	6,20	6,63	6,96	7,24	7,47	7,68	7,87	8,03
9	4,60	5,43	5,96	6,35	6,66	6,91	7,13	7,32	7,49	7,65
10	4,48	5,27	5,77	6,14	6,43	6,67	6,87	7,05	7,21	7,36
11	4,39	5,14	5,62	5,97	6,25	6,48	6,67	6,84	6,99	7,13
12	4,32	5,04	5,50	5,84	6,10	6,32	6,51	6,67	6,81	6,94
13	4,26	4,96	5,40	5,73	5,98	6,19	6,37	6,53	6,67	6,79
14	4,21	4,89	5,32	5,63	5,88	6,08	6,26	6,41	6,54	6,66
15	4,17	4,83	5,25	5,56	5,80	5,99	6,16	6,31	6,44	6,55
16	4,13	4,78	5,19	5,49	5,72	5,92	6,08	6,22	6,35	6,46
17	4,10	4,74	5,14	5,43	5,66	5,85	6,01	6,15	6,27	6,38
18	4,07	4,70	5,09	5,38	5,60	5,79	5,94	6,08	6,20	6,31
19	4,05	4,67	5,05	5,33	5,55	5,73	5,89	6,02	6,14	6,25
20	4,02	4,64	5,02	5,29	5,51	5,69	5,84	5,97	6,09	6,19
24	3,96	4,54	4,91	5,17	5,37	5,54	5,69	5,81	5,92	6,02
30	3,89	4,45	4,80	5,05	5,24	5,40	5,54	5,65	5,76	5,85
40	3,82	4,37	4,70	4,93	5,11	5,27	5,39	5,50	5,60	5,69
60	3,76	4,28	4,60	4,82	4,99	5,13	5,25	5,36	5,45	5,53
120	3,70	4,20	4,50	4,71	4,87	5,01	5,12	5,21	5,30	5,38
inf	3,64	4,12	4,40	4,60	4,76	4,88	4,99	5,08	5,16	5,23

PHÂN PHỐI KHOẢNG STUDENTIZED

$\alpha = 0{,}05$

r_2	r_1									
	2	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***
1	18,0	27,0	32,8	37,1	40,4	43,1	45,4	47,4	49,1	50,6
2	6,08	8,33	9,80	10,9	11,7	12,4	13,0	13,5	14,0	14,4
3	4,50	5,91	6,82	7,50	8,04	8,48	8,85	9,18	9,46	9,72
4	3,93	5,04	5,76	6,29	6,71	7,05	7,35	7,60	7,83	8,03
5	3,64	4,60	5,22	5,67	6,03	6,33	6,58	6,80	6,99	7,17
6	3,46	4,34	4,90	5,30	5,63	5,90	6,12	6,32	6,49	6,65
7	3,34	4,16	4,68	5,06	5,36	5,61	5,82	6,00	6,16	6,30
8	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92	6,05
9	3,20	3,95	4,41	4,76	5,02	5,24	5,43	5,59	5,74	5,87
10	3,15	3,88	4,33	4,65	4,91	5,12	5,30	5,46	5,60	5,72
11	3,11	3,82	4,26	4,57	4,82	5,03	5,20	5,35	5,49	5,61
12	3,08	3,77	4,20	4,51	4,75	4,95	5,12	5,27	5,39	5,51
13	3,06	3,73	4,15	4,45	4,69	4,88	5,05	5,19	5,32	5,43
14	3,03	3,70	4,11	4,41	4,64	4,83	4,99	5,13	5,25	5,36
15	3,01	3,67	4,08	4,37	4,59	4,78	4,94	5,08	5,20	5,31
16	3,00	3,65	4,05	4,33	4,56	4,74	4,90	5,03	5,15	5,26
17	2,98	3,63	4,02	4,30	4,52	4,70	4,86	4,99	5,11	5,21
18	2,97	3,61	4,00	4,28	4,49	4,67	4,82	4,96	5,07	5,17
19	2,96	3,59	3,98	4,25	4,47	4,65	4,79	4,92	5,04	5,14
20	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,90	5,01	5,11
24	2,92	3,53	3,90	4,17	4,37	4,54	4,68	4,81	4,92	5,01
30	2,89	3,49	3,85	4,10	4,30	4,46	4,60	4,72	4,82	4,92
40	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,63	4,73	4,82
60	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,55	4,65	4,73
120	2,80	3,36	3,68	3,92	4,10	4,24	4,36	4,47	4,56	4,64
inf	2,77	3,31	3,63	3,86	4,03	4,17	4,29	4,39	4,47	4,55

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

BẢNG PHÂN VỊ DURBIN-WATSON

$\alpha = 0{,}01$

n	k=1		k=2		k=3		k=4		k=5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0,81	1,07	0,70	1,25	0,59	1,46	0,49	1,70	0,39	1,96
16	0,84	1,09	0,74	1,25	0,63	1,44	0,53	1,66	0,44	1,90
17	0,87	1,10	0,77	1,25	0,67	1,43	0,57	1,63	0,48	1,85
18	0,90	1,12	0,80	1,26	0,71	1,42	0,61	1,60	0,52	1,80
19	0,93	1,13	0,83	1,26	0,74	1,41	0,65	1,58	0,56	1,77
20	0,95	1,15	0,86	1,27	0,77	1,41	0,68	1,57	0,60	1,74
21	0,97	1,16	0,89	1,27	0,80	1,41	0,72	1,55	0,63	1,71
22	1,00	1,17	0,91	1,28	0,83	1,40	0,75	1,54	0,66	1,69
23	1,02	1,19	0,94	1,29	0,86	1,40	0,77	1,53	0,70	1,67
24	1,04	1,20	0,96	1,30	0,88	1,41	0,80	1,53	0,72	1,66
25	1,05	1,21	0,98	1,30	0,90	1,41	0,83	1,52	0,75	1,65
26	1,07	1,22	1,00	1,31	0,93	1,41	0,85	1,52	0,78	1,64
27	1,09	1,23	1,02	1,32	0,95	1,41	0,88	1,51	0,81	1,63
28	1,10	1,24	1,04	1,32	0,97	1,41	0,90	1,51	0,83	1,62
29	1,12	1,25	1,05	1,33	0,99	1,42	0,92	1,51	0,85	1,61
30	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
31	1,15	1,27	1,08	1,34	1,02	1,42	0,96	1,51	0,90	1,60
32	1,16	1,28	1,10	1,35	1,04	1,43	0,98	1,51	0,92	1,60
33	1,17	1,29	1,11	1,36	1,05	1,43	1,00	1,51	0,94	1,59
34	1,18	1,30	1,13	1,36	1,07	1,43	1,01	1,51	0,95	1,59
35	1,19	1,31	1,14	1,37	1,08	1,44	1,03	1,51	0,97	1,59
36	1,21	1,32	1,15	1,38	1,10	1,44	1,04	1,51	0,99	1,59
37	1,22	1,32	1,16	1,38	1,11	1,45	1,06	1,51	1,00	1,59
38	1,23	1,33	1,18	1,39	1,12	1,45	1,07	1,52	1,02	1,58
39	1,24	1,34	1,19	1,39	1,14	1,45	1,09	1,52	1,03	1,58
40	1,25	1,34	1,20	1,40	1,15	1,46	1,10	1,52	1,05	1,58
45	1,29	1,38	1,24	1,42	1,20	1,48	1,16	1,53	1,11	1,58
50	1,32	1,40	1,28	1,45	1,24	1,49	1,20	1,54	1,16	1,59
55	1,36	1,43	1,32	1,47	1,28	1,51	1,25	1,55	1,21	1,59
60	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56	1,25	1,60
65	1,41	1,47	1,38	1,50	1,35	1,53	1,31	1,57	1,28	1,61
70	1,43	1,49	1,40	1,52	1,37	1,55	1,34	1,58	1,31	1,61
75	1,45	1,50	1,42	1,53	1,39	1,56	1,37	1,59	1,34	1,62
80	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,60	1,36	1,62
90	1,50	1,54	1,47	1,56	1,45	1,59	1,43	1,61	1,41	1,64
100	1,52	1,56	1,50	1,58	1,48	1,60	1,46	1,63	1,44	1,65

k: Số biến độc lập

BẢNG PHÂN VỊ DURBIN-WATSON

$\alpha = 0{,}05$

n	k=1		k=2		k=3		k=4		k=5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

k: Số biến độc lập

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

TRỢ GIÚP BÀI TẬP

Bài 3. $\bar{X} = 5{,}41$; $S^2 = 7{,}6$

Bài 4. $\bar{X} = 16$; $S^2 = 150{,}87$

Bài 5. $\bar{X} = 25{,}04$; $S^2 = 792$

Bài 6. (Tuổi nghề : $\bar{X} = 8{,}8$; $S^2 = 43{,}17$)

(Năng suất : $\bar{X} = 13{,}33$; $S^2 = 7{,}10$)

Bài 7. $\bar{X} = 2{,}85$; $S^2 = 0{,}94$

Bài 9 và 10. Giá thành = Chi phí sản suất / Sản lượng

Bài 11. Lợi suất = Lợi tức / Vốn đầu tư

Bài 14. Ngoại ô $\bar{X} = 3{,}53$; $S^2 = 1{,}86$

Ven đô $\bar{X} = 5{,}06$; $S^2 = 2{,}85$

Trung tâm $\bar{X} = 6{,}64$; $S^2 = 2{,}65$

Bài 15. Tuổi nghề:

$$\sum(x_i - \bar{x})^2 = 221 \quad , \quad \sum(x_i - \bar{x})^3 = 1058 \quad , \quad \sum(x_i - \bar{x})^4 = 16255$$

Năng suất:

$$\sum(y_i - \bar{y})^2 = 28{,}9 \quad , \quad \sum(y_i - \bar{y})^3 = -9{,}8 \quad , \quad \sum(y_i - \bar{y})^4 = 199{,}2$$

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = 54{,}9$$

Bài 16. Chi phí quảng cáo:

$$\sum(x_i - \bar{x})^2 = 991440 \quad , \quad \sum(x_i - \bar{x})^3 = 44269920$$

$$\sum(x_i - \bar{x})^4 = 70542297920$$

Tỉ suất lợi nhuận trên vốn:

$$\sum(y_i - \bar{y})^2 = 36{,}9 \quad , \quad \sum(y_i - \bar{y})^3 = 10{,}08 \quad , \quad \sum(y_i - \bar{y})^4 = 230{,}82$$

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = -122$$

Bài 22. Xem ý nghĩa vận dụng của phân phối nhị thức. Xác suất trả lời đúng một câu hoàn toàn ngẫu nhiên (không biết gì cả) là p=0,2.

Bài 23. Vận dụng định lý giới hạn trung tâm cho số trung bình của n biến ứng với n sản phẩm được chọn.

Bài 28. $\bar{X} = 5{,}1$; $S^2 = 0{,}0085$

Bài 29. $\bar{X} = 4{,}88$; $S^2 = 0{,}0152$

Bài 30. Điều kiện hợp đồng tương đương hai giả thuyết.

$\bar{X} = 9{,}75$; $S^2 = 0{,}03$

Bài 33. $\bar{X} = 2{,}52$; $S^2 = 0{,}003$

Bài 34. $\bar{X} = 10{,}34$; $S^2 = 0{,}0266$

Bài 35. d_i = BBM – BBC ($\bar{d} = 3{,}37$; $s_d^2 = 7{,}27$)

Bài 36. d_i = X – Y ($\bar{d} = 0{,}16$; $s_d^2 = 0{,}027$)

Bài 37. Máy 1 ($\bar{X} = 12{,}26$; $S^2 = 0{,}02$)

Máy 2 ($\bar{X} = 12{,}31$; $S^2 = 0{,}0167$)

Bài 38. Máy X ($\bar{X} = 5{,}07$; $S^2 = 0{,}0155$)

Máy Y ($\bar{X} = 4{,}86$; $S^2 = 0{,}0159$)

Bài 39. $\bar{X} = 8{,}2$; $S^2 = 10{,}23$, df = 77

Bài 42. $\bar{X} = 10{,}24$; $S^2 = 0{,}0082$

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Bài 58 Quảng cáo A: $\bar{x} = 8{,}17$, $s_x = 0{,}75$

x_i	z_i	$F_n(z)$	F(z)
7	-1.55	1/6	0.06
8	-0.22	2/6	0.41
8	-0.22	3/6	0.41
8	-0.22	0,67	0.41
9	1.11	0,83	0.87
9	1.11	1	0.87

Quảng cáo B: $\bar{x} = 2{,}5$, $s_x = 1{,}05$

x_i	z_i	F(z)
1	-1.43	0.08
2	-0.48	0.32
2	-0.48	0.32
3	0.48	0.68
3	0.48	0.68
4	1.43	0.92

Quảng cáo C: $\bar{x} = 4{,}67$, $s_x = 1{,}63$

x_i	z_i	F(z)
3	-1.02	0.15
3	-1.02	0.15
4	-0.41	0.34
5	0.20	0.58
6	0.82	0.79
7	1.43	0.92

Bài 59. Ký hiệu: X_i: Thu nhập , Y_i: Giá trị hóa đơn

K.hàng	X_i	Y_i	X_iY_i	X_i^2
1	2.5	0.8	2.00	6.25
2	1.8	0.5	0.90	3.24
3	3.4	1.4	4.76	11.56
…	…	…	…	…
20	7.4	3.6	26.64	54.76
Tổng	93.20	48.70	275.93	524.64

, TSS = 28,83 , RSS = 2,26

Bài 61. X_i : Thu nhập , Y_i : Giá trị hóa đơn , $X_i^* = \text{Ln}(X_i)$

KH	X_i^*	Y_i	$X_i^*Y_i$	X_i^{*2}
1	0.92	0.8	0.73	0.84
2	0.59	0.5	0.29	0.35
3	1.22	1.4	1.71	1.50
…	…	…	…	…
20	2.00	3.6	7.21	4.01
Tổng	28.77	48.70	80.48	45.42

$\sum (X_i - \bar{X})^2 = 4{,}04$, TSS = 28,83 , SSE = 26,92

Bài 62. X_i: Thu nhập , Y_i: Số lần mua hàng , $X_i^* = 1/X_i$

KH	X_i^*	Y_i	$X_i^*Y_i$	X_i^{*2}
1	0.40	2	0.80	0.16
2	0.59	3	1.76	0.35
3	1.22	4	4.90	1.50
…	…	…	…	…
20	2.00	11	22.02	4.01
Tổng	28.25	112.00	186.29	44.74

$\sum (X_i - \bar{X})^2 = 4{,}83$, TSS = 224,80 , SSE = 163,19

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Bài 63. Chọn biến Giá trị hóa đơn mua hàng làm biến phụ thuộc (được giải thích). Các biến còn lại làm biến độc lập (giải thích).

Ma trận dữ liệu:

$$X = \begin{pmatrix} 1 & 2,5 & 27 \\ 1 & 1,8 & 32 \\ 1 & 3,4 & 19 \\ 1 & 4,5 & 28 \\ 1 & 5,6 & 37 \\ 1 & 8,2 & 45 \\ 1 & 4,8 & 38 \\ \dots & \dots & \dots \\ 1 & 7,4 & 47 \end{pmatrix} \qquad Y = \begin{pmatrix} 0,8 \\ 0,5 \\ 1,4 \\ 2,2 \\ 3,0 \\ 4,5 \\ 2,5 \\ \dots \\ 3,6 \end{pmatrix}$$

Ma trận hệ số hồi quy: $\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} -0,205 \\ +0,539 \\ +0,003 \end{bmatrix}$

KH	$\hat{Y}_i$	$(Y_i - \bar{Y})^2$	$(\hat{Y}_i - \bar{Y})^2$	$(Y_i - \hat{Y}_i)^2$
1	1,22	2,67	1,47	0,18
2	0,86	3,74	2,48	0,13
3	1,68	1,07	0,56	0,08
…	…	…	…	…
20	3,92	1,36	2,22	0,11
Tổng	48,58	28,83	26,56	2,23

Ma trận hiệp phương sai:

$$Cov(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1} = \begin{pmatrix} 0.10953 & -0.00482 & -0.00198 \\ -0.00482 & 0.00151 & -0.00005 \\ -0.00198 & -0.00005 & 0.00005 \end{pmatrix}$$

Bài 64. Chọn giới tính nam làm giới tính cơ sở. Giới tính nữ được lượng hóa bằng một biến giả X_{4i}.

X_{4i} = { 1: Giới tính là nữ; 0: Giới tính không phải nữ }

KH	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	…	20
X_{4i}	0	1	0	0	1	1	0	0	1	0	1	0	0	0	1	…	0

Chọn biến Số lần mua hàng làm biến phụ thuộc (được giải thích). Các biến còn lại làm biến độc lập (giải thích).

Ma trận dữ liệu:

$$X = \begin{pmatrix} 1 & 2.5 & 27 & 0 \\ 1 & 1.8 & 32 & 1 \\ 1 & 3.4 & 19 & 0 \\ 1 & 4.5 & 28 & 0 \\ 1 & 5.6 & 37 & 1 \\ 1 & 8.2 & 45 & 1 \\ 1 & 4.8 & 38 & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 7.4 & 47 & 0 \end{pmatrix} \qquad Y = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \\ 4 \\ 10 \\ 7 \\ \dots \\ 11 \end{pmatrix}$$

Ma trận hệ số hồi quy: $\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} +0,196 \\ +1,393 \\ -0,023 \\ -0,343 \end{bmatrix}$

KH	$\hat{Y}_i$	$(Y_i - \bar{Y})^2$	$(\hat{Y}_i - \bar{Y})^2$	$(Y_i - \hat{Y}_i)^2$
1	3.06	12.96	6.45	1.12
2	1.63	6.76	15.79	1.89
3	4.50	2.56	1.21	0.25
…	…	…	…	…
20	9.43	29.16	14.66	2.47
Tổng	112.00	224.80	176.85	47.95

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Ma trận hiệp phương sai:

$$Cov(\hat{\beta}) = \hat{\sigma}^2 (X^{'} X)^{-1} = \begin{pmatrix} 2.711 & -0.134 & -0.043 & -0.374 \\ -0.134 & 0.037 & -0.001 & 0.042 \\ -0.043 & -0.001 & 0.001 & -0.003 \\ -0.374 & 0.042 & -0.003 & 0.655 \end{pmatrix}$$

Bài 76, 77, 78. Chú ý không có công thức trung bình nhân trực tiếp cho các a_i. Cần tính vòng qua trung bình nhân các t_i.

Bài 87. $\bar{x}_1 = 8{,}17$, $\bar{x}_2 = 2{,}5$, $\bar{x}_3 = 4{,}67$, $\bar{x} = 5{,}11$

TSS = 119,78 , SSB= 98,11

Bài 88. $\bar{x}_1 = 3$, $\bar{x}_2 = 7{,}43$, $\bar{x}_3 = 4{,}86$, $\bar{x} = 5{,}1$

TSS = 99,81 , SSB= 69,24

Bài 89. $\bar{x}_{t1} = 9$, $\bar{x}_{t2} = 3{,}0$, $\bar{x}_{t3} = 3{,}5$

$\bar{x}_{b1} = 4{,}67$, $\bar{x}_{b2} = 5{,}0$, $\bar{x}_{b3} = 6{,}0$, $\bar{x}_{b4} = 5$

$\bar{x} = 5{,}17$

TSS = 105,67 , SST = 88,67 , SSB = 3

Bài 90.

Lô đất không phải là một yếu tố ngoại lai được loại trừ trong thiết kế. Bởi vì, thực chất cùng một tên lô (lô số 1 chẳng hạn) nhưng là bốn lô khác nhau ở bốn nông trại khác nhau. Khác với nông trại (số 1 chẳng hạn) chỉ một nông trại duy nhất.

Vì vậy thiết kế đã cho thực chất chỉ là một thiết kế khối như sau:

TSS = 234,64 , SST = 205,84 , SSB = 4,64

Nông trại	Loại phân bón					Trung bình
	A	B	C	D	E	
1	2	8	7	10	6	6,6
2	2	6	3	9	8	5,6
3	3	6	4	11	8	6,4
4	2	6	3	9	8	5,6
5	1	7	3	12	9	6,4
Trung bình	2	6,6	4	10,2	7,8	6,12

Bài 91. $\bar{x}_{t1} = 4$, $\bar{x}_{t2} = 8{,}67$, $\bar{x}_{t3} = 4{,}67$, $\bar{x} = 5{,}78$

TSS = 53,56 , SST = 38,22 , SSC = 4,22 , SSR = 2,89

Bài 92. $\bar{x}_{t1} = 2{,}8$, $\bar{x}_{t2} = 4{,}6$, $\bar{x}_{t3} = 6{,}4$, $\bar{x}_{t4} = 7{,}4$, $\bar{x}_{t5} = 3{,}2$

$\bar{x} = 4{,}88$

TSS = 118,64 , SST = 79,44 , SSC = 31,44 , SSR = 1,04

Bài 93. $\bar{x}_{t1} = 9$, $\bar{x}_{t2} = 2{,}5$, $\bar{x}_{t3} = 6{,}5$, $\bar{x}_{t4} = 5{,}5$

$\bar{x} = 5{,}88$

TSS = 91,75 , SST = 86,75 , SSC = 0,25 , SSR = 2,25

Bài 94. Bảng dữ liệu trung bình:

Mức giảm giá (1000/kg)	Chi phí quảng cáo (%)			Trung bình
	0,6	1	1,4	
3	1.5	3.5	4.5	3.17
4	3.5	4.5	8.5	5.50
5	4.5	5.5	6.5	5.50
Trung bình	3.17	4.50	6.50	4.72

TSS =67,61 , SS(AB) = 63,11 , SSA = 33,78 , SSB = 21,78

Bài 95. Bảng dữ liệu trung bình:

Chính sách khuyến mại	Cách chiếu sáng			Trung bình
	B_1	B_1	B_1	
A_1	5.5	6.5	2.5	4.83
A_2	2	4.5	8.5	5.00
Trung bình	3.75	5.50	5.50	4.92

TSS = 147,33 , SS(AB) = 120,83 , SSA = 0,17 , SSB = 16,33

Bài 96. Một ô chỉ có một số liệu nên không nghiên cứu được tương tác giữa hai nhân tố. Phân tích phương sai tương tự thiết kế khối. Tuy nhiên ở đây cả hai yếu tố đều là yếu tố nghiên cứu chính.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Mức khuyến mãi (B)	Chi phí quảng cáo (A)			Trung bình
	0,6	1	1,4	
1	2	7	8	5.67
2	3	5	9	5.67
3	1	4	7	4.00
Trung bình	2.00	5.33	8.00	5.11

TSS = 62,89 , SSA = 54,22 , SSB = 5,56

SSE = TSS – SSA – SSB = 3,11

Tham khảo

THU THẬP VÀ NHẬP LIỆU

1. Các phương pháp thu thập dữ liệu

Tùy theo điều kiện thực tế và đặc điểm của hiện tượng nghiên cứu, người ta thường dùng một hoặc phối hợp một số phương cách thu thập dữ liệu cơ bản như: quan trắc trực tiếp, phỏng vấn, thử nghiệm, thu thập từ các nguồn có sẵn.

1.1. Phương pháp quan trắc trực tiếp: Theo phương pháp này, nhân viên điều tra phải trực tiếp tiếp xúc với đối tượng điều tra, trực tiếp tiến hành hoặc giám sát việc cân, đong, đo, đếm và sau đó tự ghi chép những thông tin thu thập được vào phiếu điều tra.

Phương pháp này cho dữ liệu có độ chính xác cao nhưng lại đòi hỏi nhiều nhân lực và thời gian. Mặt khác, trong thực tế không phải hiện tượng nào cũng cho phép cân, đong, đo, đếm trực tiếp được.

1.2. Phương pháp phỏng vấn: Theo phương pháp này, việc thu thập tài liệu phải được thực hiện thông qua quá trình hỏi-đáp giữa nhân viên điều tra và người cung cấp thông tin.

- *Phỏng vấn trực tiếp:* Nhân viên điều tra đến gặp trực tiếp người được điều tra, đặt câu hỏi, nghe trả lời và ghi vào phiếu điều tra.

Ưu điểm của phương pháp này là nhân viên có điều kiện quan sát cử chỉ, thái độ, tình cảm của đối tượng điều tra để phát hiện những nghi vấn và có biện pháp đối phó thích hợp. Mặt khác, nhân viên điều tra có thể giải thích kỹ các câu hỏi để nâng cao chất lượng dữ liệu thu thập được.

Tuy nhiên, nhược điểm của phương pháp này là tốn kém về thời gian, chi phí và đòi hỏi số lượng lớn nhân viên điều tra.

- *Phỏng vấn qua thư tín (tự ghi báo):* Cơ quan (nhân viên) điều tra gửi các phiếu điều tra đến các đối tượng cần phỏng vấn (trực tiếp hoặc theo đường bưu điện, internet), thỉnh cầu họ trả lời bằng cách tự ghi vào phiếu điều tra và gửi trả lại.

Ưu điểm của phương pháp này là dễ tổ chức, tiết kiệm chi phí và không đòi hỏi số lượng điều tra viên nhiều.

Tuy nhiên, nhược điểm của phương pháp này là khó kiểm soát độ chính xác của câu trả lời, nội dung điều tra thường hạn chế, tỉ lệ phiếu trả lời thường không cao. Để hạn chế các nhược điểm này cần chuẩn bị thật tốt bảng hỏi.

- *Phỏng vấn qua điện thoại:* Nhân viên điều tra ngồi tại bàn, quay số điện thoại cần phỏng vấn, đặt câu hỏi, nghe trả lời và ghi vào phiếu điều tra.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Phương pháp này đòi hỏi kỹ năng phỏng vấn thành thạo, phương tiện hỗ trợ thật tốt mới thực hiện có hiệu quả.

1.3. Phương pháp thử nghiệm: Bằng cách thay đổi các biến số của môi trường thử nghiệm, quan sát các phản ứng của hiện tượng thử nghiệm và đo lường các kết quả ứng với từng điều kiện môi trường thử nghiệm khác nhau.

- *Thử nghiệm hiện trường:* Việc thử nghiệm được tiến hành tại hiện trường thực của hiện tượng nghiên cứu. Phương pháp này có ưu điểm là có độ trung thực cao nhưng hạn chế khả năng thay đổi các điều kiện thử nghiệm.

- *Thử nghiệm trong phòng thí nghiệm:* Việc thử nghiệm được tiến hành trong phòng có các điều kiện môi trường mô phỏng thực tế. Những người tham gia được mời đóng vai. Phương pháp này có ưu điểm là có khả năng thay đổi các điều kiện thử nghiệm rất thuận tiện. Tuy nhiên, độ trung thực thường không cao.

1.4. Phương pháp thu thập từ các nguồn có sẵn: Dữ liệu được thu thập gián tiếp từ các chứng từ, sổ sách tài liệu của các cơ quan, đơn vị khác. Phương pháp này có ưu điểm là nhanh chóng, chi phí thấp. Tuy nhiên nhược điểm là chất lượng dữ liệu phụ thuộc vào các nguồn dữ liệu có sẵn. Nội dung điều tra đôi khi không đáp ứng đầy đủ yêu cầu nghiên cứu.

2. Sai số trong thu thập dữ liệu

Sai số trong thu thập dữ liệu là chênh lệch giữa trị số thu thập được và trị số thực tế của tiêu thức ấy.

- *Sai số trong điều tra thống kê thường được chia làm hai loại:*

+ *Sai số do đăng ký, ghi chép:* Là mọi sai số phát sinh trong quá trình điều tra do các nguyên nhân:

* Do nhân viên điều tra vô tình hay cố ý cân, đong, đo, đếm, ghi chép sai;

* Do người được phỏng vấn không hiểu đúng câu hỏi hay cố ý trả lời sai;

* Do phương tiện điều tra không chính xác.

+ *Sai số do tính chất đại biểu:* Là sai số chỉ tồn tại trong điều tra chọn mẫu do các nguyên nhân:

* Do việc chọn các đơn vị điều tra thiếu khách quan, thiếu khoa học;

* Do bản thân việc suy rộng.

- *Để hạn chế sai số cần chú ý hai biện pháp sau*:

+ Làm tốt công tác chuẩn bị điều tra ở tất cả các khâu;

+ Kiểm tra một cách toàn diện và theo suốt quá trình điều tra để phát hiện và có biện pháp khắc phục kịp thời.

3. Xây dựng bảng hỏi trong điều tra thống kê

3.1. Bảng hỏi và yêu cầu xây dựng bảng hỏi:

Bảng hỏi là hệ thống các câu hỏi được sắp xếp trên cơ sở các nguyên tắc, trình tự logic và nội dung nhất định nhằm giúp cho cuộc điều tra có thể thu thập được thông tin đáng tin cậy về hiện tượng, đáp ứng tốt nhất các yêu cầu nghiên cứu thống kê.

Bảng hỏi phải được thiết kế hài hòa, dễ đọc, có khả năng lôi kéo, duy trì sự quan tâm của người trả lời, đồng thời thuận lợi cho việc ghi chép, mã hóa, nhập liệu và kiểm tra về sau.

Mỗi câu hỏi phải rõ ràng, đơn giản, dễ hiểu và được hiểu thống nhất.

Các câu hỏi phải phù hợp với trình độ, khả năng của người trả lời để họ có thể trả lời chính xác và khách quan.

3.2. Phác thảo sơ lược dữ liệu cần thu thập

Để dữ liệu điều tra thu được đáp ứng các yêu cầu nghiên cứu, trước tiên cần phác thảo sơ lược dữ liệu như sau:

- *Xác định mục đính nghiên cứu:* Viết rõ những gì cuộc nghiên cứu nhắm đến. Điều này là khá đơn giản nhưng thường hay bị bỏ qua gây mất phương hướng, từ đó, gây lãng phí và kém hiệu quả cho cuộc điều tra nói chung.

- *Liệt kê danh sách những thông tin cần phải đo lường:* Trước hết cần xác định các khía cạnh, các vấn đề của hiện tượng nghiên cứu, các tiêu thức của mỗi khía cạnh, mỗi vấn đề. Trên cơ sở đó, liệt kê danh sách những từ, nhóm từ hay những câu hỏi sơ lược thể hiện các tiêu thức cần thu thập dữ liệu. Phải không ngừng rà soát xem chúng cung cấp những thông tin gì phục vụ mục đích nghiên cứu, đủ hay chưa, nó sẽ được sử dụng như thế nào trong quá trình phân tích về sau. Sau khi danh sách này hoàn thành phải xem liệu tất cả các khoản có thực sự cần thiết hay không, những khoản nào trùng lặp, không cần thiết cần phải bỏ đi, điều chỉnh hay thay thế. Nếu không, rất nhiều công sức, tiền bạc sẽ bị lãng phí để thu thập những thông tin vô bổ nhưng lại không đủ đáp ứng mục đích nghiên cứu.

- *Dự thảo kế hoạch phân tích dữ liệu:* Tiên liệu xem những dữ liệu điều tra được sẽ được sử dụng như thế nào, nên dùng kỹ thuật phân tích nào để mang lại ý nghĩa cho dữ liệu. Nếu dữ liệu nào không phù hợp với ý đồ phân tích ban đầu thì cần được điều chỉnh cho phù hợp.

Ví dụ, việc phác thảo sơ lược dữ liệu cho đề tài nghiên cứu "Sự lựa chọn nghề nghiệp của sinh viên", có thể được tiến hành như sau:

Mục đích nghiên cứu có thể được xác định là: Làm rõ các đặc trưng lựa chọn nghề nghiệp của sinh viên, mối liên hệ giữa các đặc điểm cá nhân, đặc điểm gia đình, đặc điểm nhân chủng và xã hội với sự lựa chọn nghề nghiệp của sinh viên.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

Các khía cạnh của hiện tượng nghiên cứu có thể được xác định:

+ Đặc điểm cá nhân;
+ Đặc điểm gia đình;
+ Đặc điểm nhân chủng và xã hội;
+ Đặc trưng nhu cầu vật chất trong chọn nghề;
+ Đặc trưng nhu cầu an toàn trong chọn nghề;
+ Đặc trưng nhu cầu giao lưu, hội nhập trong chọn nghề;
+ Đặc trưng nhu cầu được thừa nhận, kính trọng trong chọn nghề;
+ Đặc trưng nhu cầu hiểu biết, tự thể hiện trong chọn nghề.

Với đặc điểm cá nhân, có thể sử dụng các tiêu thức: Tuổi, giới tính, chiều cao, cân nặng, loại sức khỏe, điểm trung bình, xếp loại học tập, năng khiếu, sở thích, loại tính cách…

Với đặc điểm gia đình, có thể sử dụng các tiêu thức: nghề nghiệp của cha, nghề nghiệp của mẹ, thu nhập bình quân đầu người của gia đình, số thành viên trong gia đình…

Với đặc điểm nhân chủng và xã hội, có thể sử dụng các tiêu thức: dân tộc, tôn giáo, quê quán của cha mẹ, nơi thường trú…

Với đặc trưng nhu cầu vật chất trong chọn nghề, có thể sử dụng các tiêu thức: mức lương, tầm quan trọng của thu nhập, điều kiện vật chất của môi trường làm việc, môi trường sống…

Với đặc trưng nhu cầu an toàn trong chọn nghề, có thể sử dụng các tiêu thức: mức độ an toàn của ngành nghề, khu vực kinh tế, độ xa nơi làm việc, mức độ đi lại của công việc…

Với đặc trưng nhu cầu giao lưu, hội nhập trong chọn nghề, có thể sử dụng các tiêu thức: mức độ gắn bó với nơi làm việc, mức độ giao tiếp của công việc, mối quan hệ giữa gia đình và nghề nghiệp…

Với đặc trưng nhu cầu được thừa nhận, kính trọng trong chọn nghề, có thể sử dụng các tiêu thức: tầm quan trọng của chức vị, khả năng thăng tiến của nghề nghiệp, tầm quan trọng về địa vị xã hội của nghề nghiệp…

Với đặc trưng nhu cầu hiểu biết, tự thể hiện trong chọn nghề, có thể sử dụng các tiêu thức: khả năng được học tập thêm của nghề nghiệp, mức độ thách thức của nghề nghiệp, mức độ tự do của công việc, tầm quan trọng của việc được thể hiện năng lực cá nhân, mức độ mong muốn trở thành chủ doanh nghiệp riêng…

3.3. Lựa chọn dạng câu hỏi và kỹ thuật đặt câu hỏi

3.3.1. Lựa chọn câu hỏi theo vai trò

+ Câu hỏi sự kiện: Là những câu hỏi về các sự kiện hiện thực đã và đang tồn tại trong không gian và thời gian nhất định.

Ví dụ: Hiện nay anh (chị) đang làm nghề gì?

Với các sự kiện xảy ra trong quá khứ, có thể gợi ý bối cảnh lịch sử, địa điểm mà sự kiện đã xảy ra để nâng cao khả năng hồi tưởng của người trả lời. Với những câu hỏi mang tính chất riêng tư, kín đáo, tế nhị cần có cách diễn đạt tránh cho đối tượng tình huống khó xử làm ảnh hưởng đến độ chính xác của câu trả lời.

Loại câu hỏi này thường dễ trả lời, thông tin thu được thường có độ chính xác cao và ổn định.

+ Câu hỏi nhận thức: Là những câu hỏi về cảm nhận, đánh giá, quan điểm, tình cảm, mong muốn, động cơ bên trong… của người được hỏi.

Ví dụ: Anh (chị) cho biết đánh giá về chất lượng xe buýt hiện nay?

Độ chính xác của câu trả lời cho loại câu hỏi này phụ thuộc vào trình độ nhận thức, lập trường, quan điểm và khả năng diễn đạt của người trả lời.

Loại câu hỏi này đòi hỏi sự tinh tế, chính xác trong cách đặt câu, lựa chọn từ ngữ và hình thức câu hỏi.

+ Câu hỏi lọc: Là những câu hỏi có tác dụng phân chia những người tham gia trả lời thành các nhóm khác nhau, để sau đó sẽ có các câu hỏi dành riêng cho mỗi nhóm.

Ví dụ: Câu 10. Anh (chị) có hút thuốc lá không?

□ Có

□ Không (xin bỏ qua các câu hỏi từ 11 đến 14 để chuyển đến câu hỏi 15)

3.3.2. Lựa chọn câu hỏi theo biểu hiện của câu trả lời

- Câu hỏi đóng: Là dạng câu hỏi mà các phương án trả lời đã được đưa ra sẵn để người được hỏi lựa chọn. Có thể phân biệt hai loại câu hỏi đóng khác nhau:

+ Câu hỏi đóng lựa chọn (câu hỏi loại trừ): Các phương án trả lời được đưa ra có tính loại trừ nhau. Người trả lời chỉ được chọn một trong các phương án được đưa ra.

Ví dụ: Anh (chị) thuộc tôn giáo nào sau đây?

□ Công giáo

□ Phật giáo

□ Tôn giáo khác

□ Không tôn giáo

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

+ *Câu hỏi đóng tùy chọn (câu hỏi tuyển):* Các phương án trả lời được đưa ra không có tính loại trừ nhau. Người trả lời chỉ được chọn một hoặc một số trong các phương án được đưa ra.

Ví dụ: Anh (chị) đã từng làm công việc nào sau đây?

□ Giáo viên
□ Công nhân
□ Buôn bán
□ Thợ thủ công
□ Giám đốc công ty
□ Nông dân
□ Nhân viên kế toán

Câu hỏi đóng có ưu điểm là giúp người trả lời hiểu câu hỏi một cách thống nhất, tạo sự thuận tiện và dễ dàng cho việc trả lời. Mặt khác nó giúp việc xử lý dữ liệu được thuận tiện.

Hạn chế của câu hỏi đóng là gò ép người trả lời theo suy nghĩ chủ quan của người soạn thảo. Điều này hạn chế khả năng thu thập được những thông tin phong phú. Do đó khi sử dụng cần chú ý:

* Phải lường hết tất cả các phương án trả lời.

* Các phương án trả lời cần thống nhất theo một cách phân chia nào đó, tránh sự lộn xộn, gây khó cho người trả lời.

* Số lượng các phương án trả lời không nên quá nhiều nhưng cũng không nên rút gọn quá mức làm cho việc đo lường kém chuẩn xác.

- Câu hỏi mở: Là dạng câu hỏi không nêu trước các phương án trả lời mà người được hỏi tự do trả lời theo ý của họ.

Ví dụ: Anh (chị) cho biết nghề nghiệp hiện tại của anh chị?

Ưu điểm của câu hỏi mở là cho phép thu thập được những thông tin phong phú, bất ngờ làm cho việc nghiên cứu hiện tượng được sâu sắc.

Nhược điểm của câu hỏi mở là khó khăn trong quá trình thu thập dữ liệu và xử lý dữ liệu.

- Câu hỏi hỗn hợp: Đây là dạng câu hỏi kết hợp giữa câu hỏi đóng và câu hỏi mở. Vì vậy nó mang ưu nhược điểm của cả hai loại trên. Dạng câu hỏi này được sử dụng trong những trường hợp sau:

* Khi không nêu hết được các phương án trả lời, cần để người trả lời tự diễn đạt thêm.

* Khi không muốn người trả lời rơi vào tình thế khó chịu, hụt hẫng với các phương án của câu hỏi đóng.

Ví dụ: Anh (chị) thuộc tôn giáo nào sau đây?

□ Không tôn giáo
□ Công giáo
□ Phật giáo
□ Tôn giáo khác (xin ghi rõ:……)

3.3.3. Lựa chọn câu hỏi dựa trên loại thang đo: Theo mức độ tăng dần về độ chính xác của các loại thang đo, có một số dạng kỹ thuật đặt câu hỏi sau.

- Câu hỏi mở trên thang đo danh định : Đây là dạng câu hỏi có kỹ thuật đơn giản nhất. Tuy nhiên, dữ liệu khó tổng hợp nhất và ít khả năng ứng dụng các phương pháp phân tích nhất.

Ví dụ: Anh (chị) cho biết đánh giá về chất lượng xe buýt hiện nay?

- Câu hỏi đóng với nhiều mục chọn trả lời danh định: Nhiều mục trả lời danh định không chồng lẫn được tiên liệu trước cho câu hỏi. Đây là dạng câu hỏi đóng đơn giản nhất.

Ví dụ: Anh (chị) thuộc tôn giáo nào sau đây?

□ Công giáo
□ Phật giáo
□ Tôn giáo khác
□ Không tôn giáo

- Câu hỏi đóng với hai mục chọn trả lời đối nhau: Thường dùng các cụm từ đối nhau như: đồng ý – không đồng ý, đúng – sai, phải – không… Ví dụ: Theo bạn việc làm tốt nhất là việc làm có thu nhập cao nhất.

☐ Đồng ý
☐ Không đồng ý

- Câu hỏi đóng với nhiều mục chọn trả lời thứ bậc: Người ta thường dùng các cấp so sánh của tính từ kết hợp với một cặp từ đối nghịch để tạo thứ bậc cho các mục chọn được liệt kê sẵn bên dưới câu hỏi. Người được phỏng vấn được yêu cầu đánh dấu vào một mục mà mình quyết định chọn. Ví dụ:

Theo anh chị chương trình truyền hình là:

☐ Rất hấp dẫn
☐ Khá hấp dẫn
☐ Bình thường *(đối xứng với 5 thứ bậc)*
☐ Khá buồn tẻ
☐ Rất buồn tẻ

- Câu hỏi đóng sắp hạng thứ tự các mục trả lời: Các mục trả lời thường là các nguyên nhân, các nhu cầu, các công việc, các bộ phận, các sở thích, các thành phần, các thành tố… được liệt kê sẵn. Người trả lời được yêu cầu sắp xếp

Dit document is beschikbaar op studeersnel
Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

chúng theo các loại thứ tự như: theo tầm quan trọng, theo mức độ ưa thích, theo giá trị, theo mức độ cấp bách, theo thời gian đòi hỏi, theo mức độ cần thiết… Ví dụ: Anh chị hãy sắp xếp các yếu tố sau đây theo tầm quan trọng của chúng đối với doanh nghiệp hiện nay từ 1 đến 7 ?

☐ Vốn
☐ Công nghệ
☐ Tính sáng tạo
☐ Tinh thần dám kinh doanh, dám chịu rủi ro
☐ Sự hỗ trợ của chính quyền thành phố
☐ Lao động kỹ thuật

- Câu hỏi đóng trên thang điểm đều 5, 10 hay 100: Người được hỏi được yêu cầu cho điểm trên thang điểm 5, 10 hay 100 giữa hai thái cực của hiện tượng. Ví dụ: Anh chị hãy cho điểm từ 0 (hoàn toàn không ưa thích) đến 10 (hoàn toàn ưa thích) đối với dịch vụ xe buýt hiện nay.

- Câu hỏi đóng trên thang đo Likert: Câu hỏi điển hình là: "Xin vui lòng đọc kỹ những phát biểu sau đây. Sau mỗi câu phát biểu, hãy khoanh tròn trả lời thể hiện đúng nhất quan điểm của anh (chị).

Ví dụ: Theo anh chị tìm được việc làm hiện nay là rất khó khăn:

1	2	3	4	5
Rất không đồng ý	Có phần không đồng ý	Lưỡng lự	Có phần đồng ý	Rất đồng ý

Hoặc:

1	2	3	4	5	6	7
Hoàn toàn không đồng ý	Nói chung không đồng ý	Có chút không đồng ý	Lưỡng lự	Có chút đồng ý	Nói chung là đồng ý	Hoàn toàn đồng ý

Hầu hết các thang đo Likert có số lượng lẻ các lựa chọn trả lời như 3, 5 hoặc 7. Mục đích là để đưa ra cho người trả lời các lựa chọn trả lời có điểm giữa trung lập. Tuy nhiên, số lựa chọn trả lời chẵn buộc người trả lời phải xác định một quan điểm rõ ràng trong khi số lựa chọn trả lời lẻ cho phép họ lựa chọn dung hòa hơn. Do đó, không thể khẳng định cách nào hay hơn. Tùy trường hợp cụ thể mà có lựa chọn hợp lý.

- Câu hỏi mở trên thang đo tỉ lệ : Loại câu hỏi này chỉ thực hiện được trên một số hiện tượng nhất định. Tuy nhiên đây là loại câu hỏi cho dữ liệu có khả năng ứng dụng được nhiều nhất các phương pháp thống kê.

Ví dụ : Anh (chị) cho biết thu nhập một tháng của anh chị là bao nhiêu ?

Để mở rộng khả năng vận dụng các phương pháp thống kê, cần sử dụng tối đa có thể được các dạng câu hỏi trên các thang đo có độ chính xác cao hơn trong đo lường.

3.4. Lựa chọn từ ngữ và cách hành văn

Từ ngữ và cách hành văn cần được lựa chọn kỹ càng nhằm nâng cao chất lượng của cuộc điều tra. Khi lựa chọn từ ngữ và cách hành văn cần lưu ý những vấn đề sau:

- *Từ ngữ phải đơn giản, dễ hiểu* đối với mọi người ở mọi vùng, mọi trình độ đang được điều tra, tránh dùng từ lóng, từ chuyên môn;
- *Tránh sự diễn đạt dài dòng dễ gây lẫn lộn và khó chịu* cho người được phỏng vấn;
- *Từ ngữ phải rõ ràng, chính xác*; nếu từ đa nghĩa, chung chung sẽ dẫn đến việc thu được những dữ liệu không phù hợp hoặc vô nghĩa, ví dụ: Bạn thường dùng ***loại*** bút bi gì ?;
- *Tránh dùng từ bao nghĩa trong một câu hỏi*, nếu trong một câu hỏi có từ "và" thì nên tách thành hai câu, ví dụ: Theo anh chị dịch vụ xe buýt hiện nay có ***tốt và tiện lợi*** không ?;
- *Tránh đặt câu có tính gợi ý*, điều này dễ dẫn đến câu trả lời thiên lệch, ví dụ: ***Rất nhiều người cho rằng*** nhịp sống hiện nay là căng thẳng, anh chị có nghĩ vậy không ?;
- *Tránh đặt câu có tính định kiến* như dùng từ nhạy cảm hay đưa ra các mục lựa chọn trả lời không đối xứng, điều này dễ dẫn đến câu trả lời thiên lệch, ví dụ: Anh chị có cho rằng sinh viên hiện nay quá ***thực dụng*** ? Anh chị cho rằng điều kiện học tập của sinh viên hiện nay là:

☐ *Tuyệt vời*
☐ *Rất tốt*
☐ *Tốt*
☐ *Khá tốt*
☐ *Trên trung bình*

- *Hạn chế câu hỏi mà thông tin trả lời đòi hỏi sự tổng hợp phức tạp* từ nhiều yếu tố hay trong một thời gian dài, điều này dễ dẫn đến sự phỏng đoán cho xong chuyện của người trả lời; trường hợp này nên chia tách thành các câu hỏi theo từng yếu tố hay hỏi trong khoảng thời gian ngắn nhất có thể được, ví dụ không

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

nên hỏi: Anh chị thường đi xe buýt bao nhiêu lần ***trong năm*** ?, mà nên hỏi: Anh chị thường đi xe buýt bao nhiêu lần ***trong tuần*** (hay trong tháng) ?.

- *Nếu cần ngụy trang mục đích chính của cuộc nghiên cứu* nhằm tránh sự trả lời thiên lệch thì điều này cũng cần được lưu ý khi lựa chọn từ ngữ trong suốt bản câu hỏi.

3.5. Cấu trúc bảng hỏi phỏng vấn

Cấu trúc chung của một bảng hỏi phỏng vấn có hiệu quả thường bao gồm các phần như sau:

- *Phần lời ngỏ:* Gồm những lời tự giới thiệu về nhân viên điều tra, về đơn vị nghiên cứu, về mục đích cuộc nghiên cứu (có thể đã được ngụy trang một cách tinh vi nhưng không vi phạm nguyên tắc đạo đức), lời kêu gọi sự hợp tác, giúp đỡ.

- *Phần câu hỏi mở đầu:* Gồm những câu hỏi bao quát nhất, có tính gợi chuyện, dễ trả lời và không mang tính cá nhân nhằm làm quen và tạo thiện cảm ở người được phỏng vấn.

- *Phần câu hỏi phân hóa:* Gồm những câu hỏi giúp xác định người được phỏng vấn thuộc loại có đặc tính nghiên cứu nào để sử dụng các câu hỏi tiếp theo cho phù hợp. Phần này có hay không tùy thuộc vào cuộc nghiên cứu cụ thể.

- *Phần câu hỏi khơi gợi sự hồi tưởng:* Bao gồm các câu hỏi về các sự kiện, về các mốc thời gian, không gian, bối cảnh các sự việc, diễn tiến các sự việc, về các mối liên hệ… giúp cho người trả lời liên tưởng và dần dần nhớ lại các sự việc trong quá khứ, nhờ đó việc trả lời các câu hỏi tiếp theo được thuận tiện và chính xác hơn.

- *Phần câu hỏi trọng tâm:* Đây là phần chủ yếu, gồm những câu hỏi đi vào chiều sâu, có tính thông tin cao phục vụ mục tiêu nghiên cứu (có thể đã được ngụy trang).

- *Phần thông tin xác thực:* Gồm các câu hỏi thể hiện phép lịch sự, sự quan tâm và các thông tin xác nhận cuộc phỏng vấn như: họ và tên, địa chỉ, nghề nghiệp, tuổi tác, giới tính, chữ ký của người được phỏng vấn, họ tên người phỏng vấn, ngày, giờ phỏng vấn…

Hình thức bản câu hỏi phải lịch sự, trang trọng và ngắn gọn nhất có thể được.

4. Kiểm tra và hiệu chỉnh dữ liệu

***4.1. Kiểm tra, đánh giá dữ liệu*:** Kiểm tra đánh giá dữ liệu trước khi sử dụng là một công việc không thể thiếu. Việc sử dụng vội vàng, thiếu chặt chẽ dựa trên các tài liệu điều tra một cách dễ dãi thì kết quả nghiên cứu là không đáng tin cậy, thậm chí rất có hại.

- Tiến hành xem xét một cách kỹ lưỡng phương án điều tra để bảo đảm tính khoa học và chất lượng của dữ liệu.

- Nghiên cứu kỹ các bảng hỏi và những chỉ dẫn về các thủ tục phỏng vấn để phát hiện ra những khả năng có thể dẫn đến các sai sót.

- Nếu là tài liệu điều tra chọn mẫu cần kiểm tra qui trình lấy mẫu theo phương án điều tra và trên thực địa để bảo đảm tính ngẫu nhiên và tính đại diện của mẫu.

- Nếu dữ liệu thứ cấp thì đánh giá dữ liệu theo mức độ phù hợp, mức độ lỗi thời của dữ liệu, uy tín của đơn vị cung cấp dữ liệu. Nếu cần, kiểm tra thông tin về phương pháp thu thập dữ liệu, mục đích sử dụng dữ liệu của đơn vị cung cấp.

Nếu dữ liệu không đáng tin cậy thì không nên sử dụng mà phải yêu cầu điều tra lại.

***4.2. Hiệu chỉnh dữ liệu*:** Với các phiếu điều tra có vấn đề cần phải được hiệu chỉnh. Phương pháp hiệu chỉnh là quay trở lại thẩm vấn nhân viên điều tra hoặc người được điều tra để xác minh hoặc suy luận từ manh mối của các câu trả lời khác, các phiếu điều tra khác.

- Những phiếu điều tra có những sai sót thuộc về ghi chép, ngôn từ hoặc chữ viết không rõ ràng thì có thể xem xét để hiệu chỉnh lại, loại bỏ một số câu hay toàn bộ phiếu.

- Phát hiện những phiếu điều tra giả tạo do nhân viên điều tra phịa ra. Những phiếu điều tra này thường có các câu trả lời dị thường rất dễ bị phát hiện. Cần phải điều tra lại hay loại bỏ những phiếu này.

- Những phiếu điều tra không đầy đủ, có những câu bị bỏ trống, trả lời nửa chừng hay không rõ ý, lựa chọn nhiều phương án trong câu hỏi loại trừ. Cần xem có thể hiệu chỉnh hay loại bỏ một số câu hay toàn bộ phiếu.

- Phát hiện những phiếu trả lời thiếu nhất quán, không hợp logic. Ví dụ Tình trạng hôn nhân là có gia đình. Trong khi đó ở câu khác tuổi mới 15. Cần xem có thể hiệu chỉnh hay loại bỏ một số câu hay toàn bộ phiếu.

- Phát hiện những phiếu trả lời cung cấp những thông tin không liên quan đến câu hỏi, có thể do người trả lời hiểu nhầm hoặc không thuộc đối tượng điều tra của câu hỏi đó. Cần xem có thể hiệu chỉnh hay loại bỏ một số câu hay toàn bộ phiếu.

Việc loại bỏ một số câu hay một số phiếu điều tra chỉ thực hiện được với với dữ liệu mẫu và phải rất hạn chế vì việc làm này sẽ ảnh hưởng đến kết quả nghiên cứu. Với dữ liệu tổng thể thì phải điều tra lại những phiếu này.

Dit document is beschikbaar op studeersnel

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)

5. Nhập liệu và mã hóa dữ liệu trên SPSS:

5.1. Với câu hỏi mở, định lượng:

Câu 1. Bạn cho biết thu nhập một tháng là bao nhiêu ?.....

→ *Dữ liệu câu hỏi này được nhập theo một biến và không cần mã hóa.*

5.2. Với câu hỏi đóng, danh định, được chọn một đáp án:

Câu 2. Anh (chị) thuộc tôn giáo nào sau đây?

□ Công giáo

□ Phật giáo

□ Tôn giáo khác

□ Không tôn giáo

→ *Dữ liệu câu hỏi này được nhập theo một biến C2 và được mã hóa như sau:*

Tên biến	Kiểu biến	Nhãn biến	Mã hóa dữ liệu
C2	String	Tôn giáo	1 = Công giáo 2 = Phật giáo 3 = Tôn giáo khác 4 = Không tôn giáo

5.3. Với câu hỏi đóng, danh định, được chọn nhiều đáp án:

Câu 3. Anh (chị) đã từng làm những công việc nào sau đây?

□ Giáo viên

□ Công nhân

□ Buôn bán

→ *Dữ liệu dạng câu hỏi này được nhập theo k biến với k là số đáp án của câu hỏi và được mã hóa như sau:*

Tên biến	Kiểu biến	Nhãn biến	Mã hóa dữ liệu
C3.1	Numeric	Giáo viên	1 = Giáo viên 0 = Chưa làm gv
C3.2	Numeric	Công nhân	1 = Công nhân 0 = Chưa làm cn
C3.3	Numeric	Buôn bán	1 = Buôn bán 0 = Chưa làm bb

Chú ý kiểu biến ở Câu 3 đúng ra là kiểu string. Tuy nhiên ở đây chọn kiểu numeric với mục đích dễ tổng hợp ba biến C3.1, C3.2 và C3.3.

5.4. Với câu hỏi đóng, định lượng, được chọn một đáp án:

Câu 4. Bạn cho biết chi tiêu một tháng là bao nhiêu ?

□ Dưới 1 triệu

□ 1 – 2 triệu

□ 2 – 3 triệu

□ Trên 3 triệu

→ Dữ liệu câu hỏi này có thể được nhập tương tự câu 1, không cần mã hóa mà nhập liệu dựa trên các trị số giữa. Câu hỏi này cũng có thể nhập liệu như câu 2, dựa trên các giá trị mã hóa cho từng khoảng chi tiêu.

5.5. Với câu hỏi mở, danh định:

Câu 5. Anh (chị) cho biết nghề nghiệp hiện tại?.....

→ Dữ liệu câu hỏi này được nhập theo một biến và có thể được mã hóa như sau: Cần xem xét tất cả các biểu hiện trả lời trên khoảng 30% số phiếu để quyết định phân chia chúng vào bao nhiêu nhóm là thích hợp. Trường hợp này cần ứng dụng lý thuyết phân tổ (được trình bày rõ ở chương 2) để quyết định phân chia chúng làm bao nhiêu tổ là vừa. Sau đó, mã hóa theo từng tổ:

Tên biến	Kiểu biến	Nhãn biến	Mã hóa dữ liệu
C4	String	Nghề nghiệp	1 = Tổ 1 (Công nhân) 2 = Tổ 2 (Công chức) 3 = Tổ 3 (Nông dân) 4 = Tổ 4 (Doanh nhân) 5 = Tổ 5 (Nghề khác)

Trong quá trình nhập liệu nếu xuất hiện thêm nhiều biểu hiện "Nghề khác" nữa, nếu cần, có thể tách tổ 5 thêm một số tổ mới và mã hóa thêm.

5.6. Với câu hỏi hỗn hợp:

Ví dụ: Câu 6. Anh (chị) thuộc tôn giáo nào sau đây?

□ Không tôn giáo

□ Công giáo

□ Phật giáo

□ Tôn giáo khác (xin ghi rõ:…….)

→ Dữ liệu câu hỏi này được nhập theo một biến và được mã hóa tương tự như câu 2. Riêng đáp án "Tôn giáo khác" được mã hóa như "Nghề khác" trong câu 5.

5.7. Với câu hỏi đóng trên thang đo Likert *(xem trang 142)**:***

→ Dữ liệu câu hỏi này có thể được nhập theo một biến tương tự câu 1.

Downloaded by Th? Nguy?n (nguyenthu238756@gmail.com)